

1/69

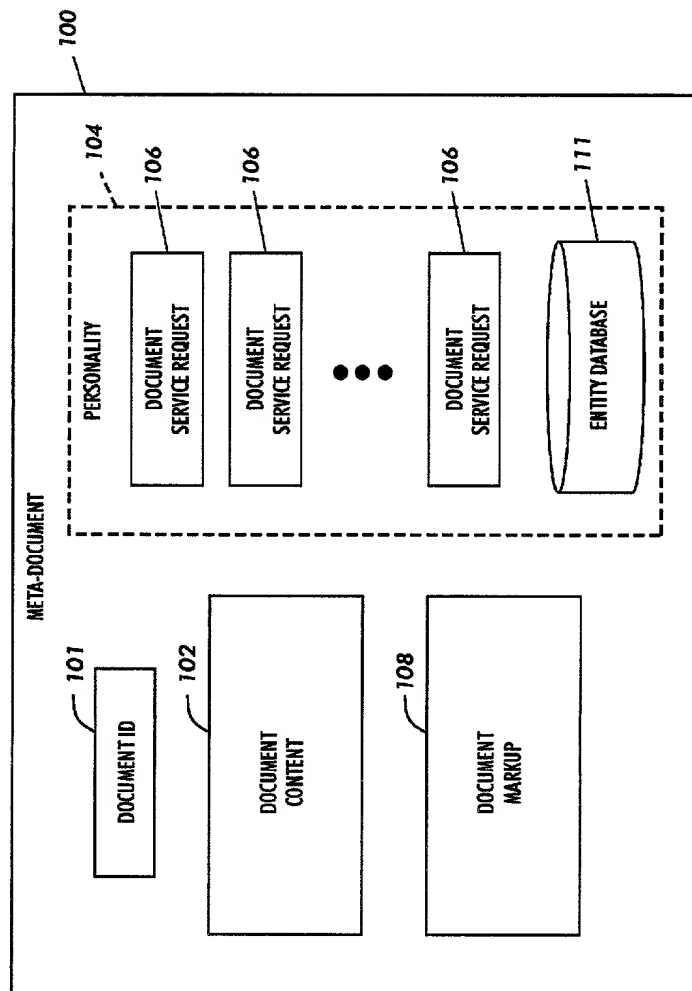


FIG. 1

2/69

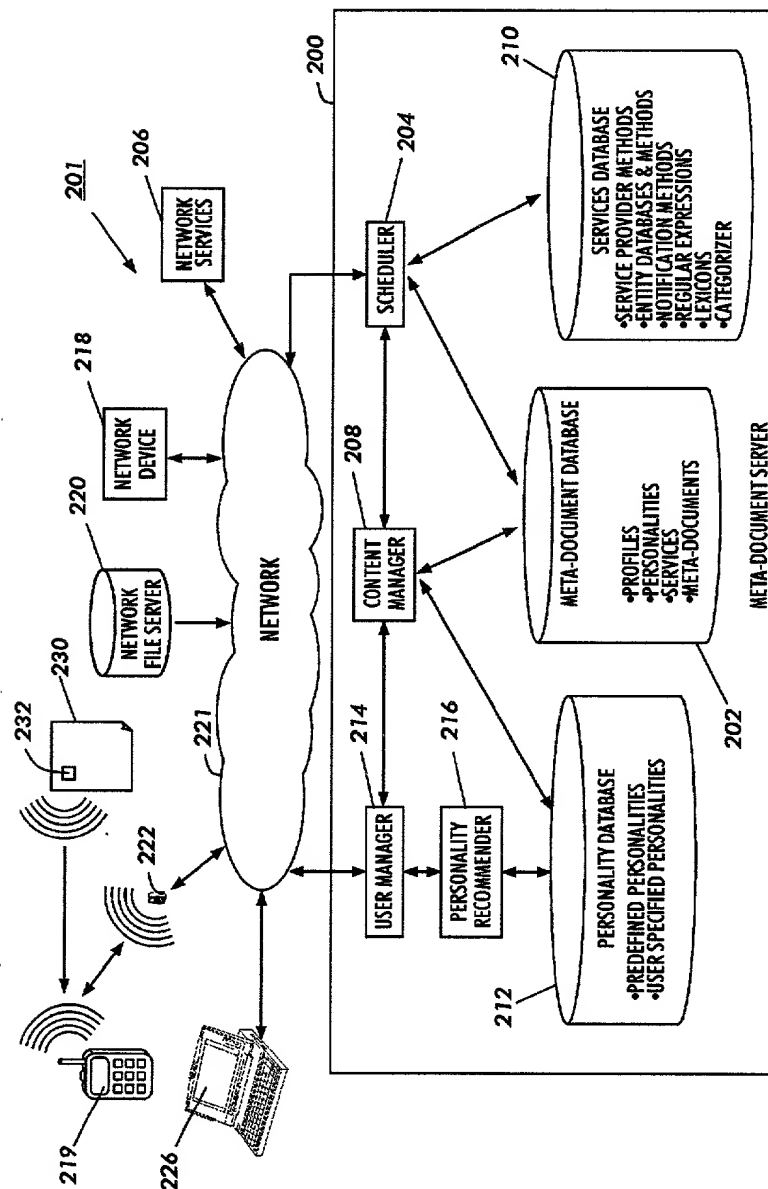


FIG. 2

3/69

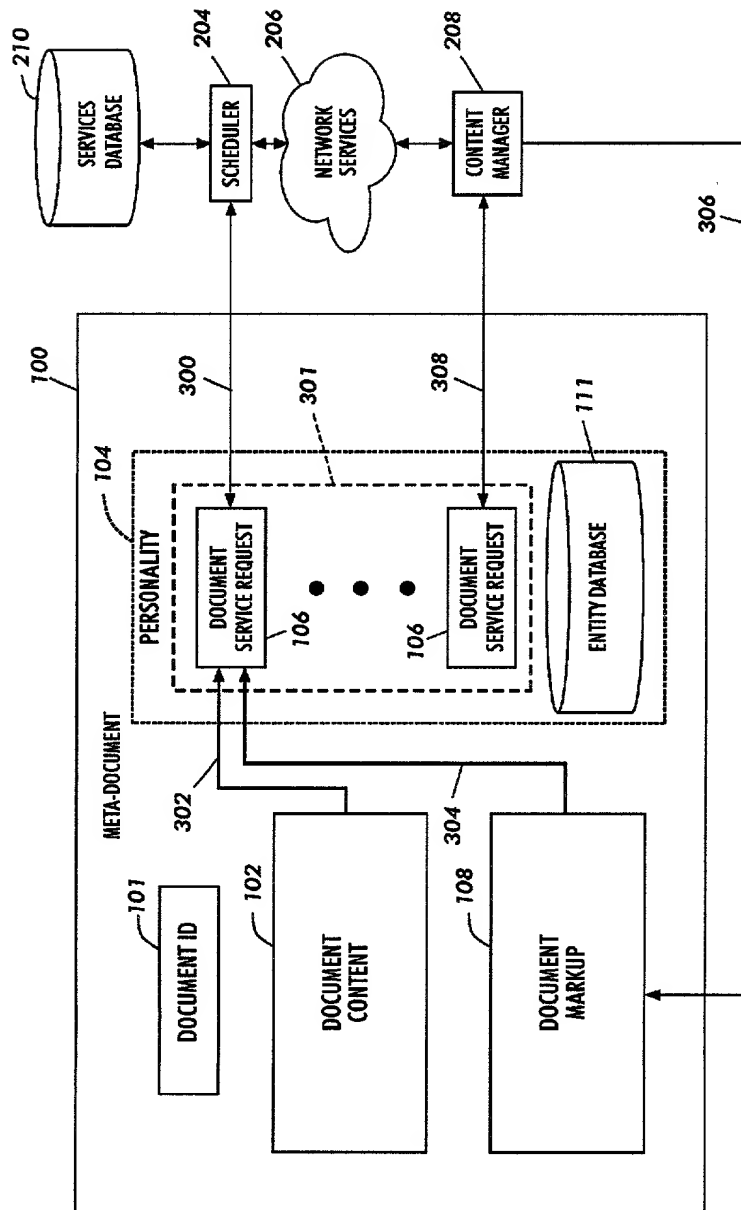


FIG. 3

4/69

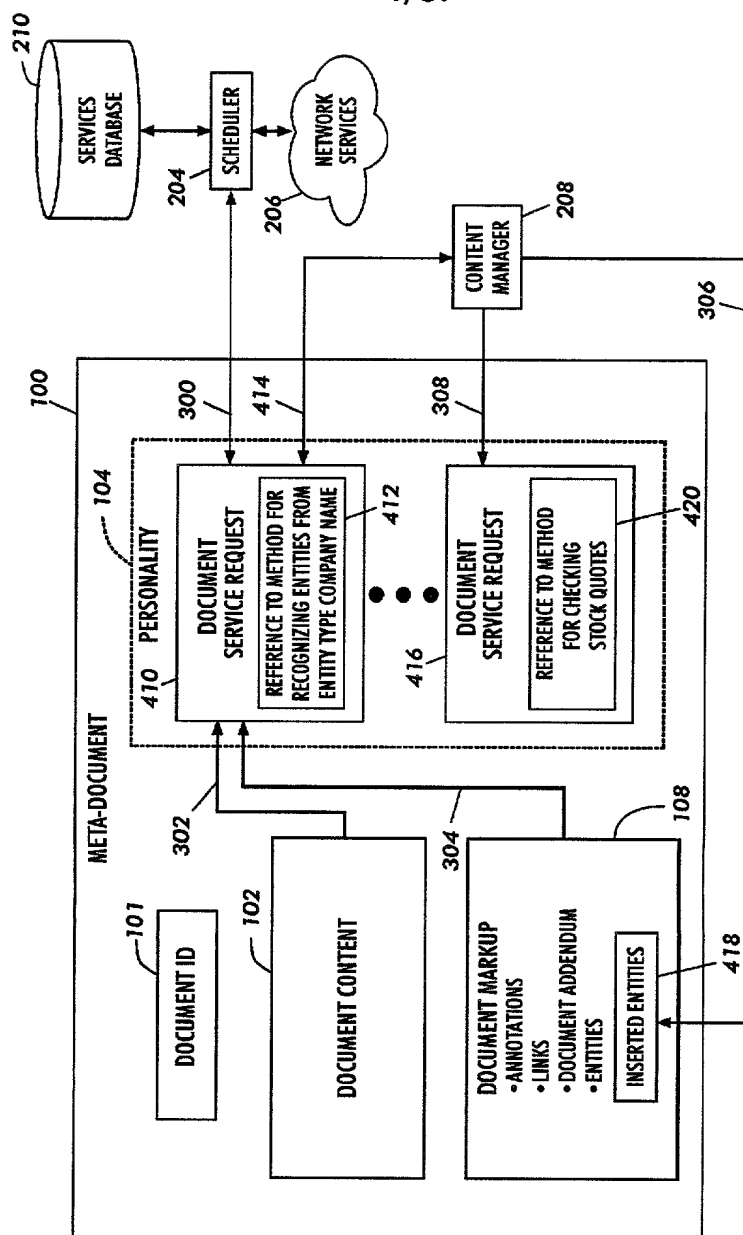


FIG. 4

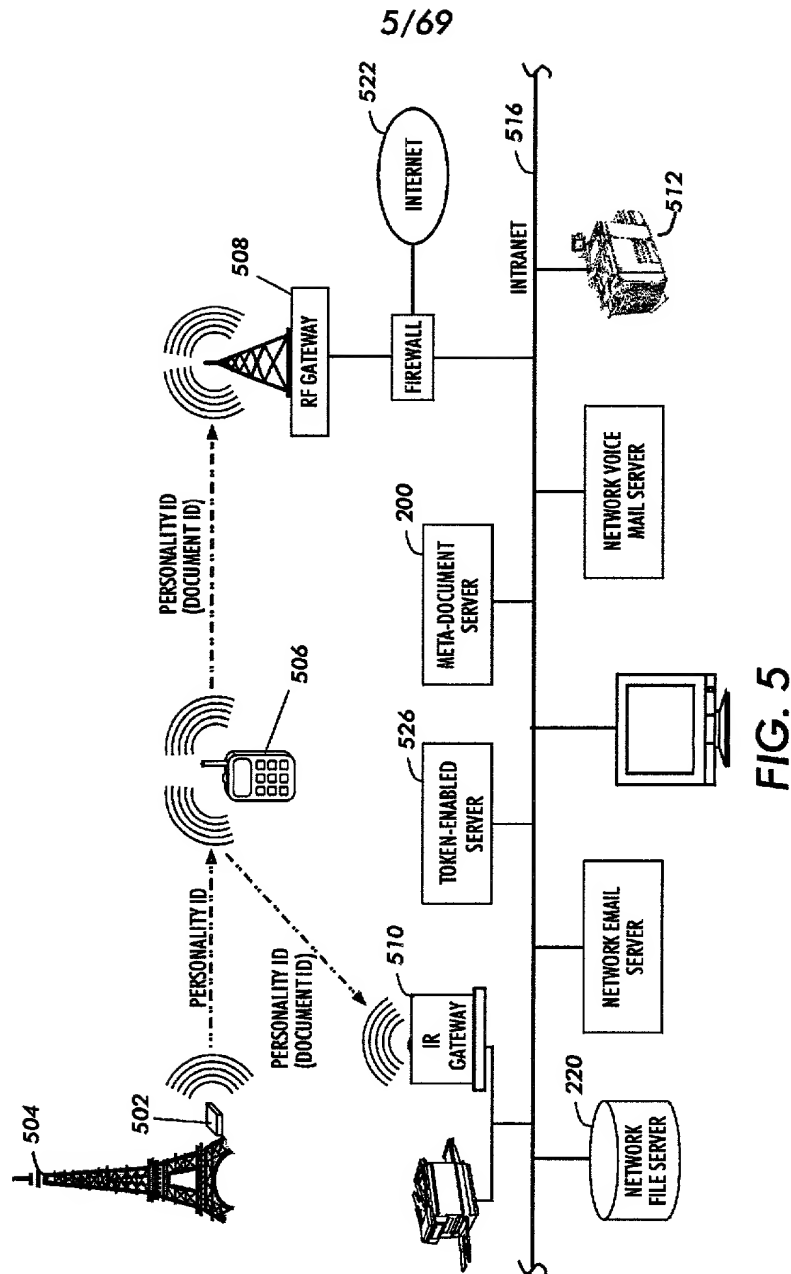
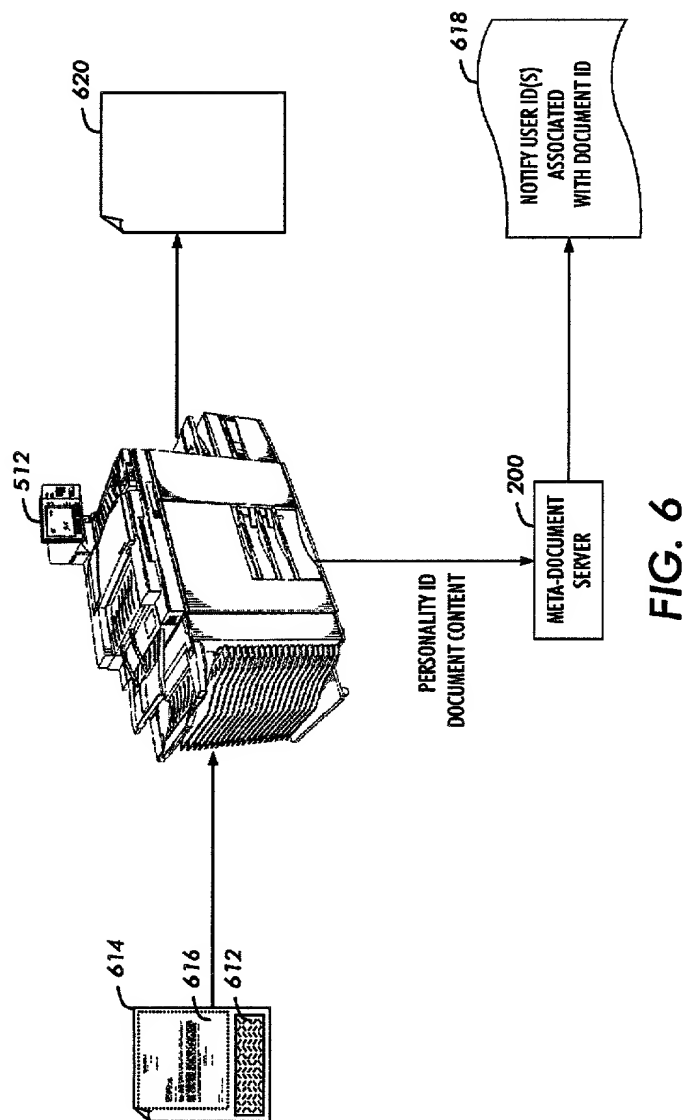


FIG. 5

6/69



7/69

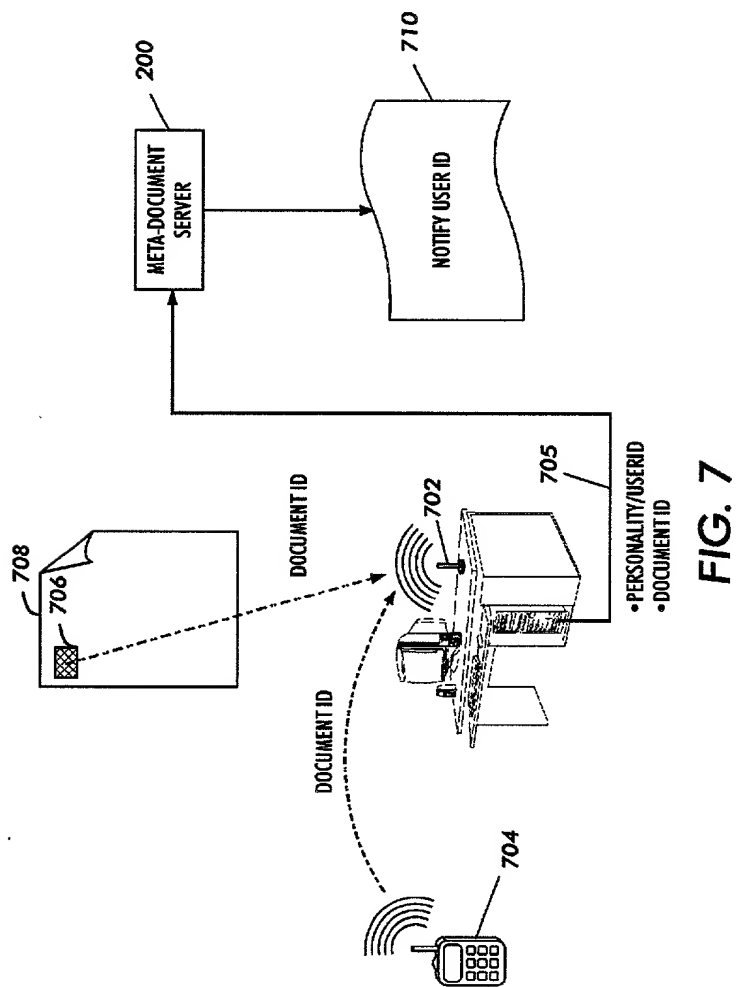


FIG. 7

8/69

**PRINT**

**PRINTER**  
 NAME:  **PROPERTIES**  
 STATUS: IDLE  
 TYPE: XEROX 4517 POSTSCRIPT LEVEL 2  
 WHERE: TIMES.GRENOBLE.XRCE.XEROX.COM:PASSTHRU  
 COMMENT: XEROX N2125 B&W DUPLEX - LE CHATEAU - 2ND FLOOR

**PAGE RANGE**  
☒ ALL  
☐ CURRENT PAGE ☐ SELECTION  
☐ PAGES:   
 ENTER PAGE NUMBERS AND/OR PAGE RANGES SEPERATED BY COMMAS. FOR EXAMPLE 1,3,5-12

**COPIES**  
 NUMBER OF COPIES:   
☒ COLLATE

**ENRICHMENT**  
☒ PERSONALITY:  ☐ SELECT A PERSONALITY FOR ME  
 APPLY ENRICHMENT TO:  
☒ ALL  
☐ CURRENT PAGE ☐ SELECTION  
☐ PAGES:   
 INSERT ENRICHMENT AS:  
☒ LINKS  
☐ CONTENT **OPTIONS...**

**PRINT WHAT:**  **ZOOM**    
**PAGES PER SHEET:**   
**PRINT:**  **SCALE TO PAPER SIZE:**   
**OPTIONS...** **OK** **CANCEL**

FIG. 8



9/69

PRINT DOCUMENT PROPERTIES

LAYOUT PAPER/QUALITY ENRICHMENT

902 PERSONALITY: BUSINESS SELECT PERSONALITY FOR ME 900

PROVIDE ENRICHMENT BY: 901

PRINTING WITH DOCUMENT  
STORING AT SERVER

INSERT ENRICHMENT AS:

LINKS  
CONTENT

WHEN STORING ENRICHMENT AT SERVER: 904

EMAIL ME WHEN COMPLETED  
ADDRESS:   
DO NOT NOTIFY ME

CAPTURE POINT: 908 906

AT WORKSTATION  
AT PRINT SERVER

ADVANCED

OK CANCEL

FIG. 9



FIG. 10

11/69

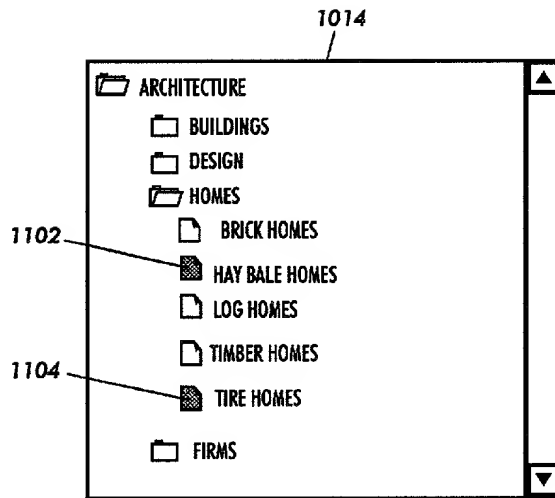


FIG. 11

12/69

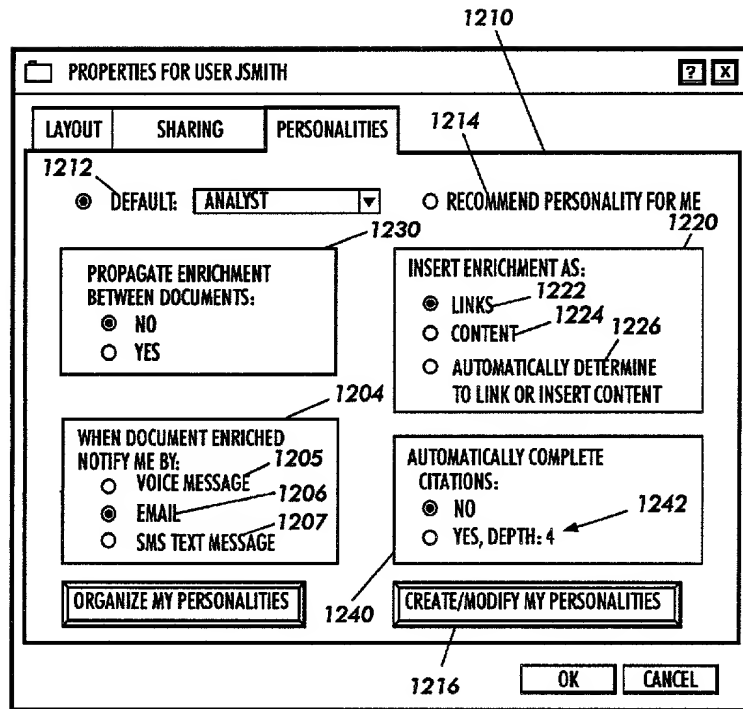


FIG. 12

13/69

1310

☐ CREATE/MODIFY A GROUP OF EXISTING PERSONALITIES

1316

1312

Name	Creation Date
Technology Watch in Computer Sciences	2001-03-16 14:08:06
Watch Business	2001-03-16 14:08:08
Watch Legal	2001-03-16 14:08:10
Watch Intellectual Property	2001-03-16 14:08:10
Scientific Watch in Medicine	2001-03-16 14:08:10

1314

1318

Click on this icon to view all available service categories

Search Engine

- *luckygoogle People Google Lucky Result*

Stock Quotes

- *Bloomberg Stock Quotes*

OK CANCEL

FIG. 13

14/69

1400	1404	1422	1420	1408	1426	1412	1414	1430	1442	1440	1444	1450
SEARCH ENGINE	NAME	ENTITY TYPE	ANNOTATE	DEPTH	PROVIDER	COST	QUERY	REFRESH				
<input checked="" type="checkbox"/>	COMPUTING	COMPUTING	ALWAYS	0	YAHOO	FREE	LINK	DAILY	1452			
<input type="checkbox"/>	BIOLOGY	BIOLOGY	ALWAYS	0	GOOGLELUCKY	NULL	LINK	NEVER	1451			
<input type="checkbox"/>	MEDICINE	MEDICINE	ALWAYS	0	ALTAVISTA	FEE	LINK	AUTO	1455			
<input type="checkbox"/>	APPLY PAIRS OF ENTITIES AND IDENTIFY 3RD PARTY ENTITIES	COMPANIES	EXPERT	0	GOOGLELUCKY	NULL	CONTENT	NULL	1453			
<input type="checkbox"/>	APPLY PAIRS OF ENTITIES USING GOOGLELUCKY	PRODUCTS	ALWAYS	2	NORTHERNLIGHTS	NULL	LINK	WEEKLY	1456			
<input type="checkbox"/>	COUNTRIES	COUNTRIES	MOVIE	0	GOOGLELUCKY	FEE	CONTENT	UNTIL 2003	1454			
<input type="checkbox"/>	ASTRONOMY	ASTRONOMY	ALWAYS	1	GOOGLE	NULL	LINK	MONTHLY				
<input checked="" type="checkbox"/>	PEOPLE	PEOPLE	ALWAYS	3	GOOGLE PEOPLE	NULL	LINK	NULL				

VIEW HIERARCHY  
SET DEPTH 1432

FIG. 14

15/69

1502

The screenshot shows a software window titled "CREATE/MODIFY MY PERSONALITIES" with a standard Windows-style title bar containing a question mark and a close button. The window contains the following elements:

- A "NAME:" label followed by a "BROWSE" button and a text input field (labeled 1504).
- A radio button (labeled 1506) next to the text "CREATE/MODIFY A GROUP OF EXISTING PERSONALITIES".
- A radio button (labeled 1508) next to a box containing the text "USE SELECTED DOCUMENTS AND/OR FOLDERS TO CREATE A NEW PERSONALITY". Inside this box is a "BROWSE" button and a text input field (labeled 1509).
- A radio button (labeled 1510) next to a box containing the text "USE SELECTED FILE/WEBSITE TO CREATE A NEW PERSONALITY". Inside this box is a "BROWSE" button and a text input field.
- A radio button (labeled 1520) next to a box containing the text "CREATE QUESTION FORM USING PERSONALITY LEXICONS". Inside this box is a "QUESTION:" label and a text input field.
- A "Hint:" label followed by a text input field (labeled 1514).
- A "KNOWLEDGE LEVEL:" label with two radio buttons: "NOVICE" (selected) and "EXPERT" (labeled 1516).
- "OK" and "CANCEL" buttons (labeled 1512).

FIG. 15

16/69

1600

CREATE/MODIFY A GROUP OF EXISTING PERSONALITIES

NAME: TECHNOLOGY WATCH IN ORGANIC CHEMISTRY

OPERATION:

☒ MERGE 1602

☐ SUBTRACT 1604

PERSONALITY A ▼

WITH/FROM

PERSONALITY B ▼

OK CANCEL

FIG. 16



17/69

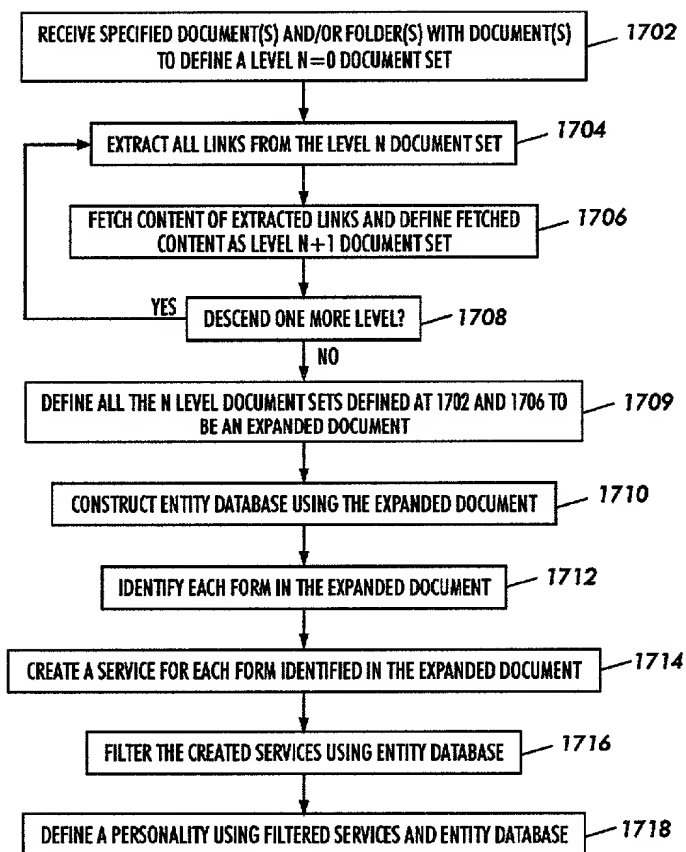


FIG. 17

18/69

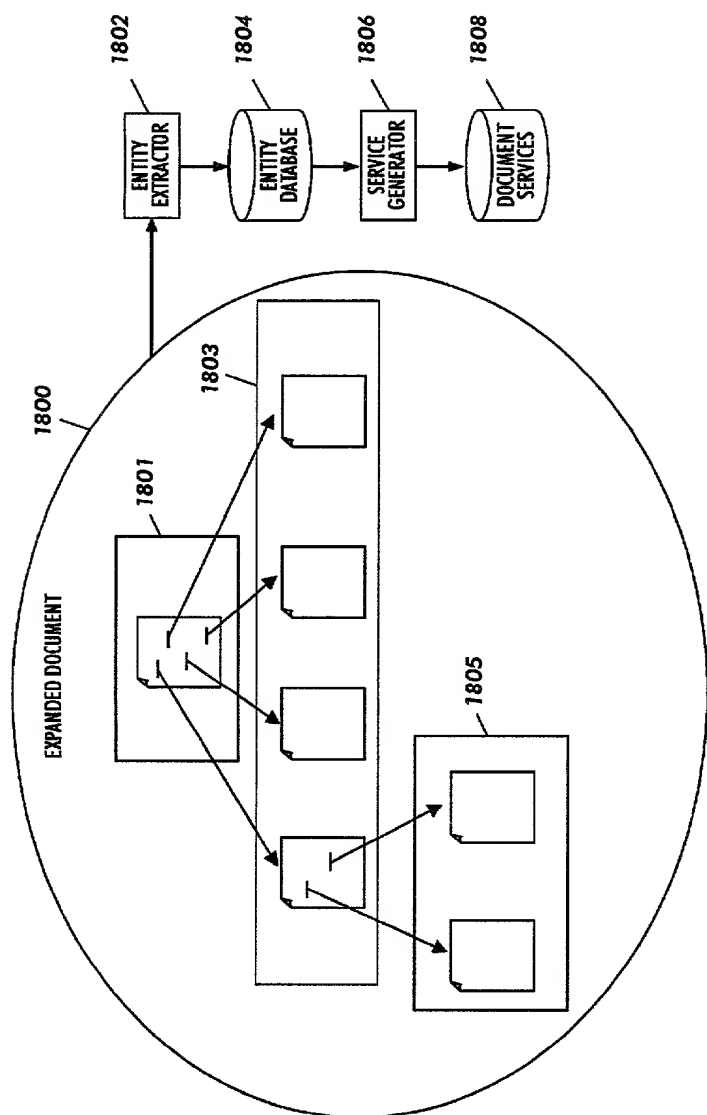


FIG. 18

19/69

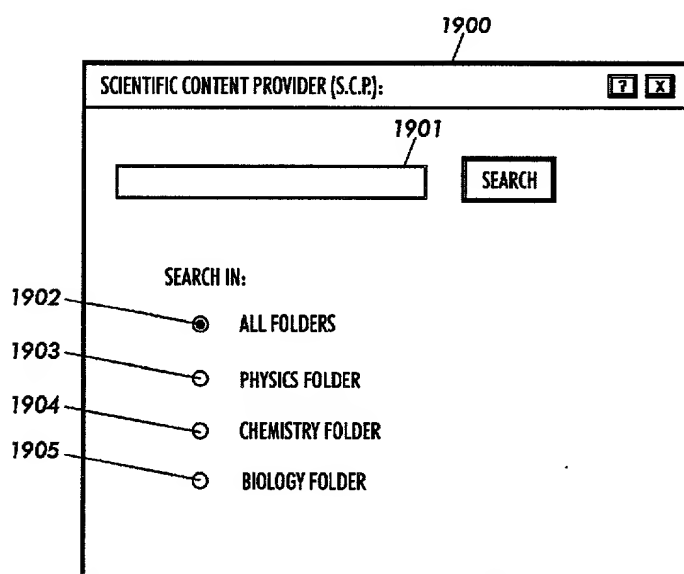



FIG. 19

20/69

SCIENTIFIC SEARCH 									
SEL	NAME	ENTITY TYPE	ANNOTATE	DEPTH	PROVIDER	COST	QUERY	REFRESH	
<input checked="" type="checkbox"/>	ALL FOLDERS RESULT OF SCIENTIFIC CONTENT PROVIDER	STANDARD	ALWAYS	0	S.C.P.	FREE	CONTENT	NULL	
<input type="checkbox"/>	PHYSICS RESULT OF SCIENTIFIC CONTENT PROVIDER	PHYSICS	ALWAYS	0	S.C.P.	FREE	CONTENT	NULL	
<input type="checkbox"/>	CHEMISTRY RESULT OF SCIENTIFIC CONTENT PROVIDER	CHEMISTRY	ALWAYS	0	S.C.P.	FREE	CONTENT	NULL	
<input type="checkbox"/>	BIOLOGY RESULT OF SCIENTIFIC CONTENT PROVIDER	BIOLOGY	ALWAYS	0	S.C.P.	FREE	CONTENT	NULL	

2004

2002

FIG. 20

FIG. 20 is a screenshot of a web browser displaying a search results page for scientific content. The page shows a table with columns for selection, name, entity type, annotate, depth, provider, cost, query, and refresh. The table lists four search results: 'ALL FOLDERS RESULT OF SCIENTIFIC CONTENT PROVIDER', 'PHYSICS RESULT OF SCIENTIFIC CONTENT PROVIDER', 'CHEMISTRY RESULT OF SCIENTIFIC CONTENT PROVIDER', and 'BIOLOGY RESULT OF SCIENTIFIC CONTENT PROVIDER'. The first result is selected, and the 'QUERY' column for all results is 'CONTENT'. The 'REFRESH' column for all results is 'NULL'.

21/69

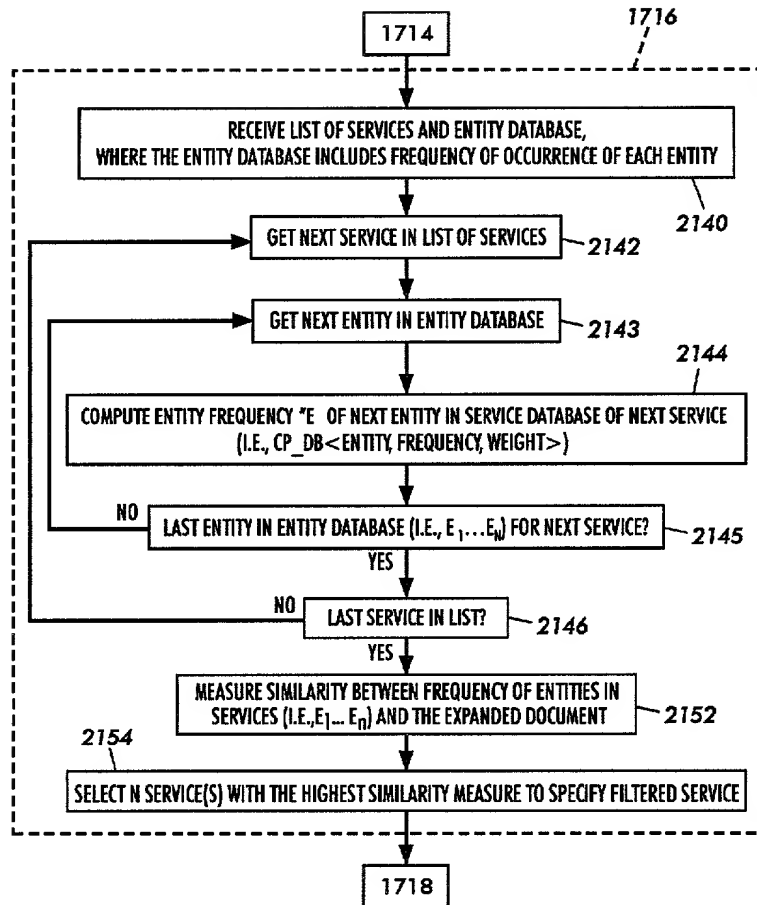
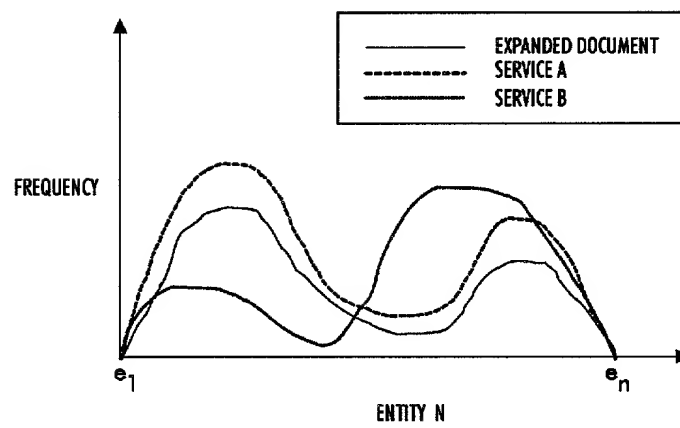


FIG. 21

22/69

**FIG. 22**

23/69

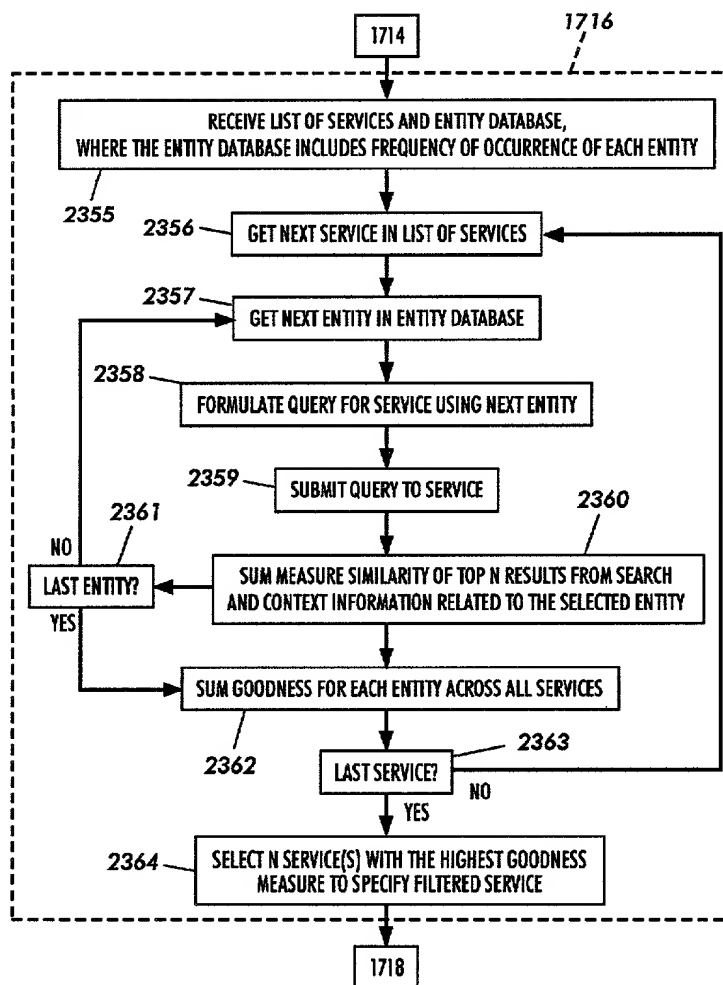


FIG. 23

24/69

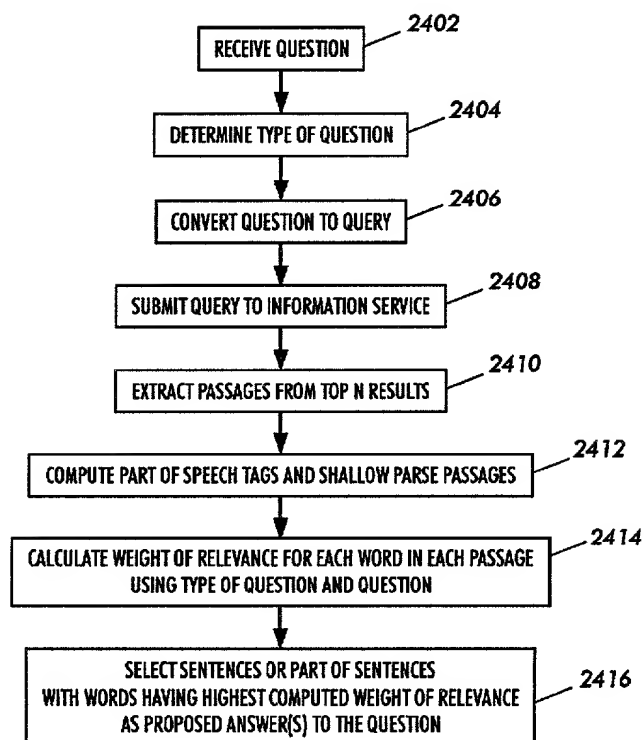


FIG. 24



25/69

2502

E-LEARNING SERVICES ASSOCIATED WITH E-LEARNING PERSONALITY	
SEL	
<input checked="" type="checkbox"/>	LINK WORDS OR MULTWORD EXPRESSIONS FOUND IN THE DOCUMENT TO THEIR DEFINITIONS
<input type="checkbox"/>	LINK WORDS OR MULTWORD EXPRESSIONS FOUND IN THE DOCUMENT TO THEIR TRANSLATIONS
<input type="checkbox"/>	LINK EACH TEXT UNIT (I.E., DOCUMENT, PARAGRAPH, PHRASE, WORD) TO A TUTORIAL CONCERNING EACH TEXT UNIT
<input type="checkbox"/>	LINK EACH TEXT UNIT (I.E., DOCUMENT, PARAGRAPH, PHRASE, WORD) TO AN EXPERT CONCERNING IDENTIFIED SUBJECTS
<input type="checkbox"/>	LINK EACH TEXT UNIT (I.E., DOCUMENT, PARAGRAPH, PHRASE, WORD) TO AN INTERACTIVE COURSE ABOUT IDENTIFIED SUBJECTS
<input checked="" type="checkbox"/>	LINK EACH TEXT UNIT (I.E., DOCUMENT, PARAGRAPH, PHRASE, WORD) TO A LIST OF REAL WORLD CLASSES ABOUT IDENTIFIED SUBJECTS
<input checked="" type="checkbox"/>	LINK TEXT UNITS TO (I.E., DOCUMENT, PARAGRAPH, PHRASE, WORD) LEARNING RESOURCES CONCERNING EACH TEXT UNITS

2504  
2506  
2508  
2510  
2512  
2514  
2516

FIG. 25

26/69

2602

LANGUAGE LEARNING SERVICES ASSOCIATED WITH LANGUAGE LEARNING PERSONALITY	
SEL	
<input checked="" type="checkbox"/>	LINK WORDS OR MULTWORD EXPRESSIONS FOUND IN THE DOCUMENT TO THEIR DEFINITIONS
<input type="checkbox"/>	LINK WORDS OR MULTWORD EXPRESSIONS FOUND IN THE DOCUMENT TO THEIR TRANSLATIONS
<input type="checkbox"/>	LINK EACH SENTENCE, OR PHRASE, TO A GRAMMATICAL DESCRIPTION OF THE STRUCTURE OF THE SENTENCE OR PHRASE
<input type="checkbox"/>	LINK EACH WORD, MULTWORD EXPRESSION, PHRASE OR SENTENCE TO OTHER INSTANCES OF THE SAME IN DIFFERENT CONTEXTS FROM THE PRESENT
<input type="checkbox"/>	LINK EACH WORD, MULTWORD EXPRESSION, PHRASE OR SENTENCE TO ONE OR MORE INTERACTIVE GRAMMAR EXERCISES CONCERNING THAT ELEMENT
<input checked="" type="checkbox"/>	LINK EACH TEXT UNIT (I.E., DOCUMENT, PARAGRAPH, PHRASE, WORD) TO A LIST OF REAL WORLD CLASSES ABOUT IDENTIFIED LANGUAGE
<input checked="" type="checkbox"/>	LINK TEXT UNITS TO (I.E., DOCUMENT, PARAGRAPH, PHRASE, WORD) LANGUAGE LEARNING RESOURCES CONCERNING EACH TEXT UNITS

2604

2606

2608

2610

2612

2614

2616

FIG. 26

FIG. 26 is a screenshot of a user interface for language learning services. The interface displays a list of services associated with a language learning personality. The services are listed in a table with columns for a selection checkbox and a description. The services include linking words or multiword expressions to their definitions, linking words or multiword expressions to their translations, linking each sentence or phrase to a grammatical description of the structure of the sentence or phrase, linking each word, multiword expression, phrase or sentence to other instances of the same in different contexts from the present, linking each word, multiword expression, phrase or sentence to one or more interactive grammar exercises concerning that element, linking each text unit (i.e., document, paragraph, phrase, word) to a list of real world classes about identified language, and linking text units to (i.e., document, paragraph, phrase, word) language learning resources concerning each text unit. The selection checkboxes for the first, second, third, fourth, fifth, sixth, and seventh services are checked, while the checkbox for the eighth service is unchecked.

27/69

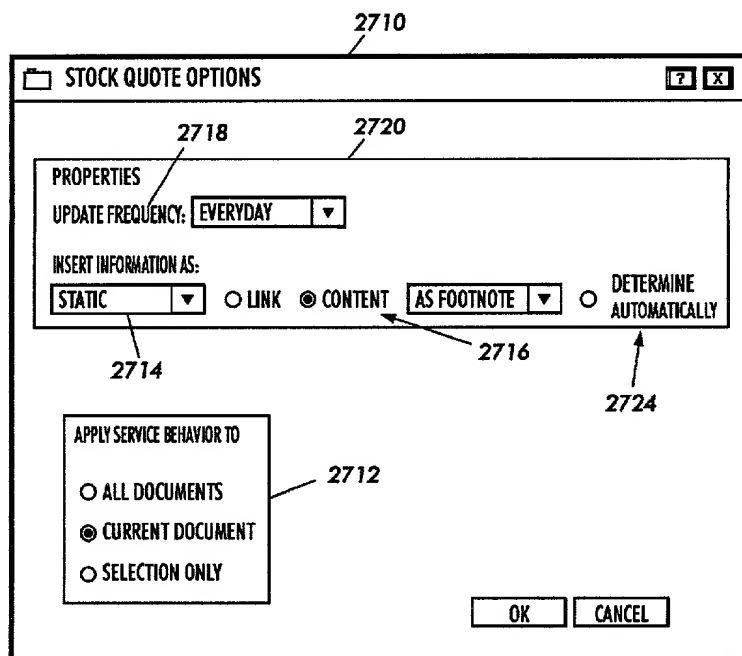


FIG. 27

28/69

2800

ANNOTATE PROPERTY

ANNOTATE ENTITIES WHEN:

2802 ☐ ALWAYS

2804 ☒ EXPERT MODE:

2806 ☐ NOVICE MODE:

2808 CHANGE REFERENCE

2801

2805

2803

2807

2810

2811

AND

ANNOTATE ONLY THOSE ENTITIES THAT OCCUR WITH A FREQUENCY LESS THAN: 4 IN THE REFERENCE DOCUMENT(S) AND/OR DATABASE(S) AT WWW.XRCE.XEROX.COM/DB/EXPERT.

ANNOTATE ONLY THOSE ENTITIES THAT OCCUR WITH A FREQUENCY GREATER THAN: 20 IN THE REFERENCE DOCUMENT(S) AND/OR DATABASE(S) AT WWW.XRCE.XEROX.COM/DB/NOVICE.

ANNOTATE ENTITY WHEN IT APPEARS IN DOCUMENT CONTENT WITH A FREQUENCY GREATER THAN: 4.

OK CANCEL

FIG. 28

29/69

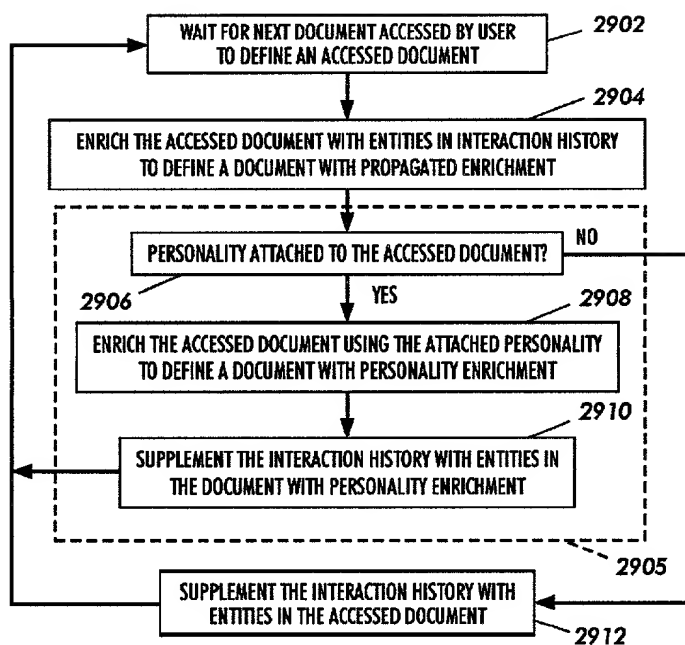


FIG. 29

30/69

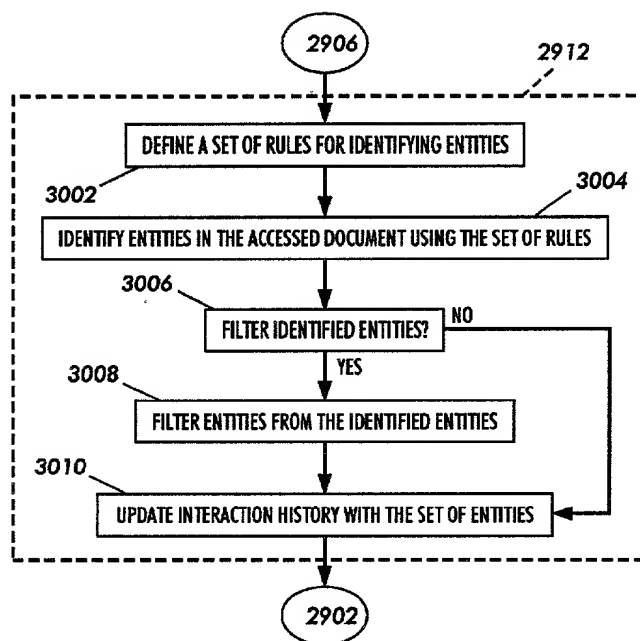


FIG. 30

31/69

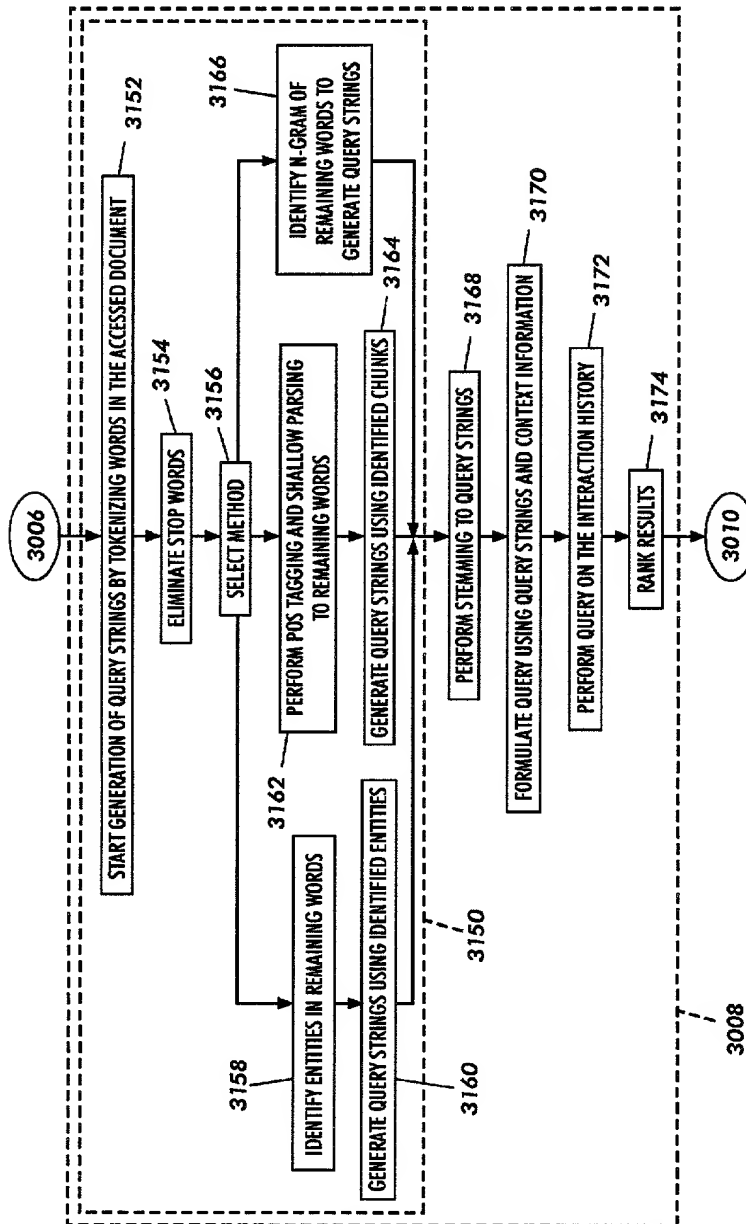


FIG. 31

32/69

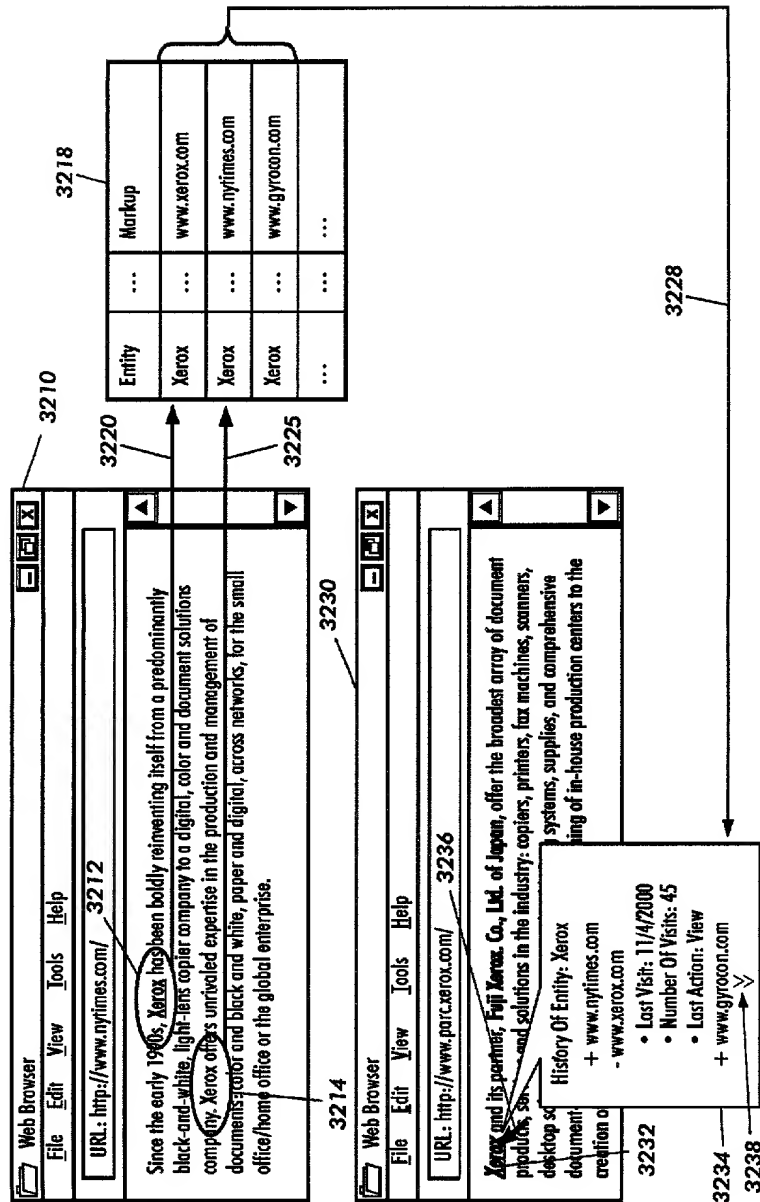


FIG. 32



33/69

3218

ENTITY	ENTITY TYPE	VISITS	LAST VISIT	ACTION LIST	DMOZ CATEGORY	LOCATION	POS	MARKUP
XEROX	BUSINESS	45	11/4/2000	VIEW (43), PRINT(2)	BUSINESS	40	NOUN	WWW.XEROX.COM
XEROX	BUSINESS	1	5/1/2001	VIEW(1)	FINANCE	689	NOUN	WWW.NYTIMES.COM
XEROX	BUSINESS	4	2/22/2001	EMAIL(1), VIEW(3)	COMPUTER	343	NOUN	WWW.GYROCON.COM
...	...	...	...	...	...	...	...	...

FIG. 33

34/69

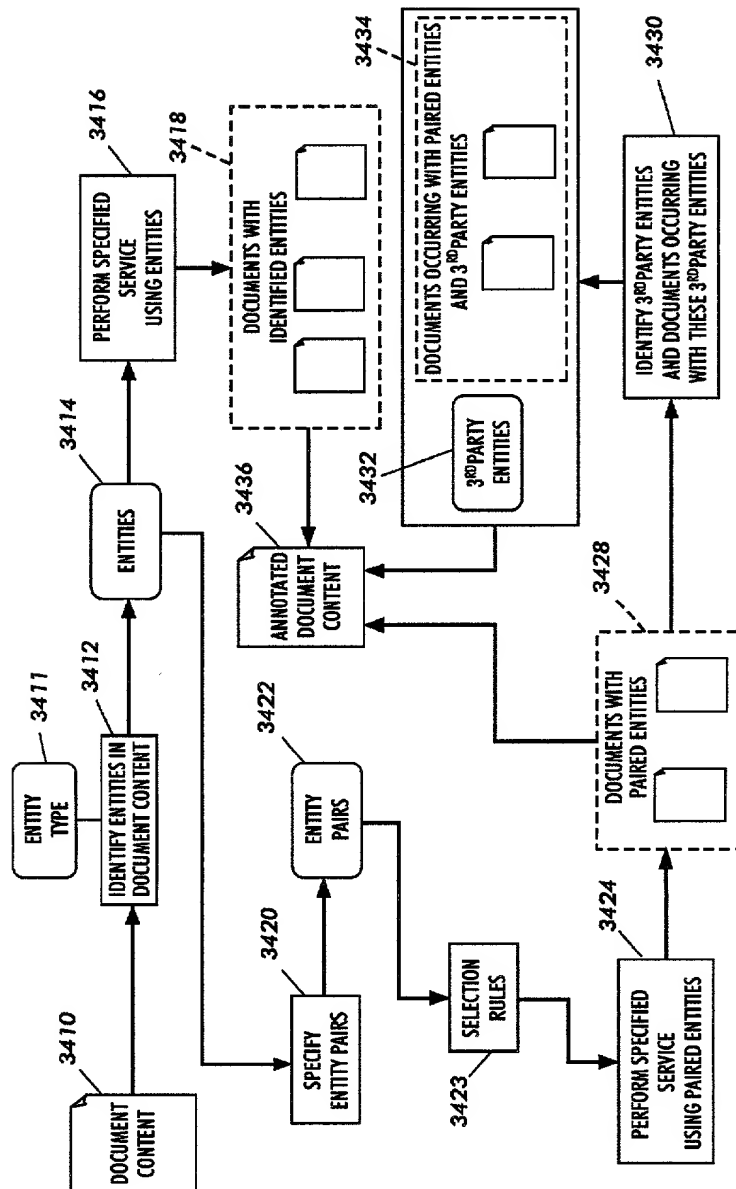
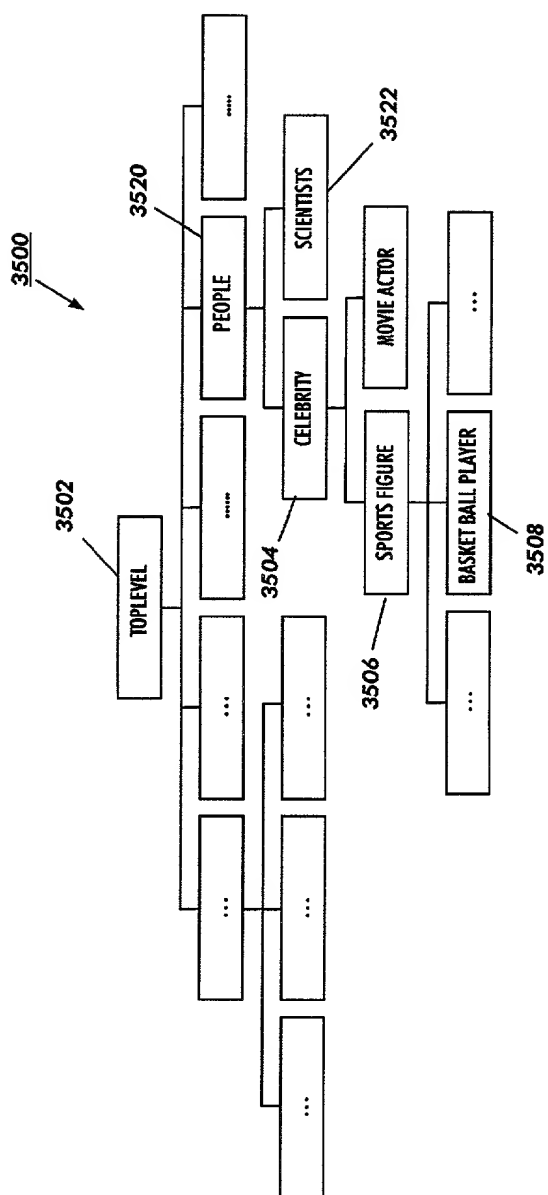


FIG. 34

35/69



**FIG. 35**

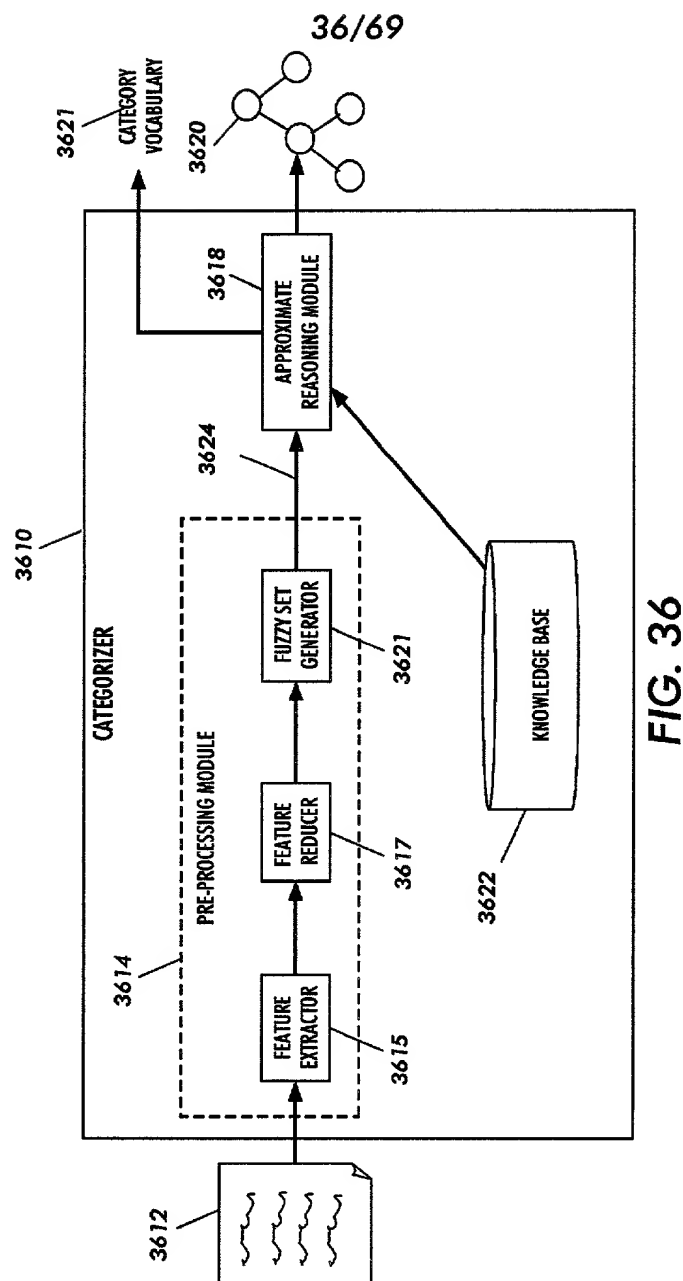


FIG. 36

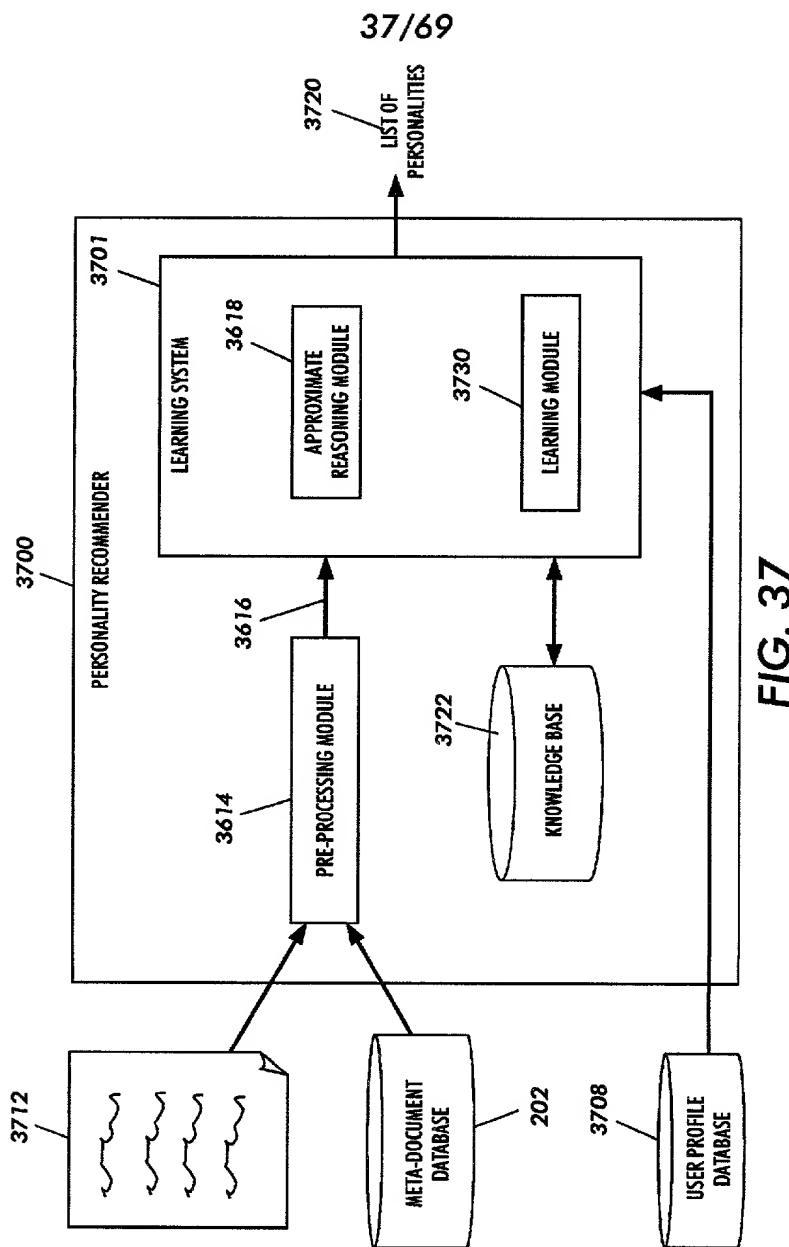


FIG. 37

38/69

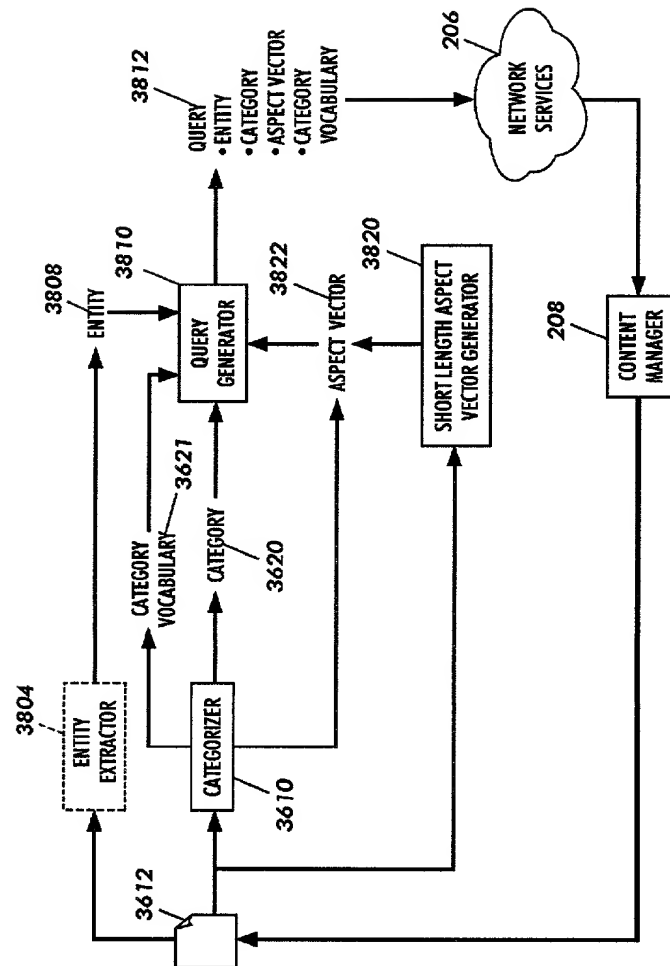


FIG. 38

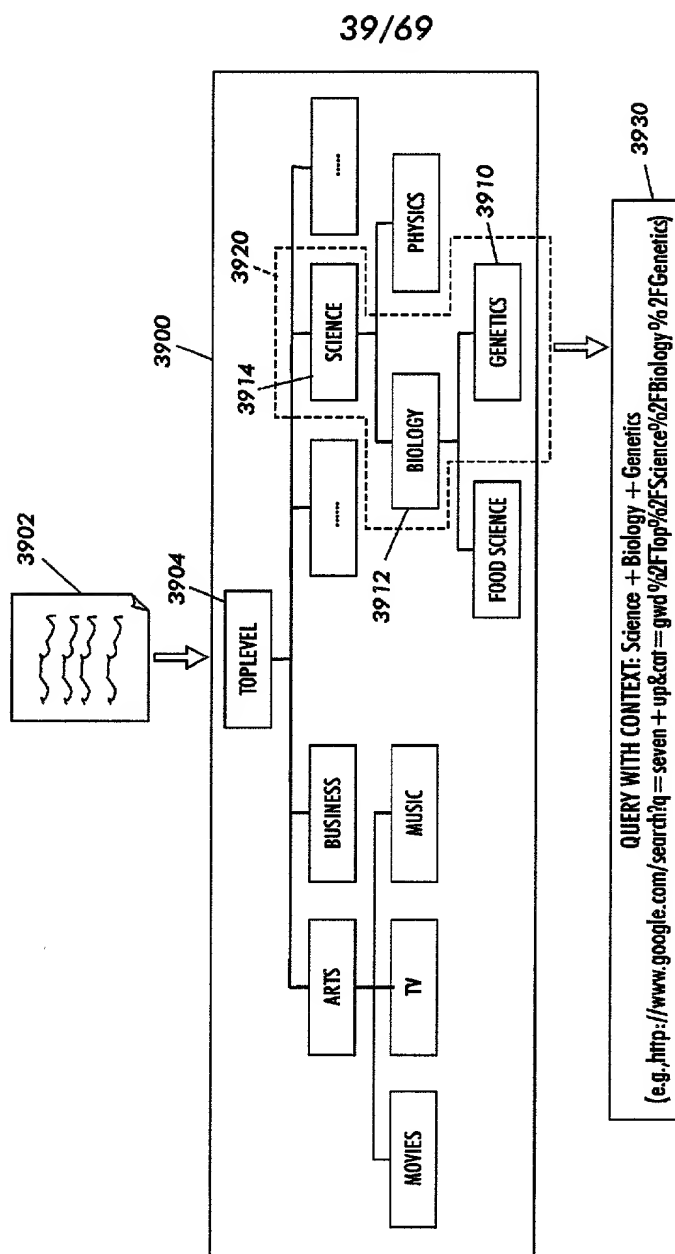


FIG. 39

40/69

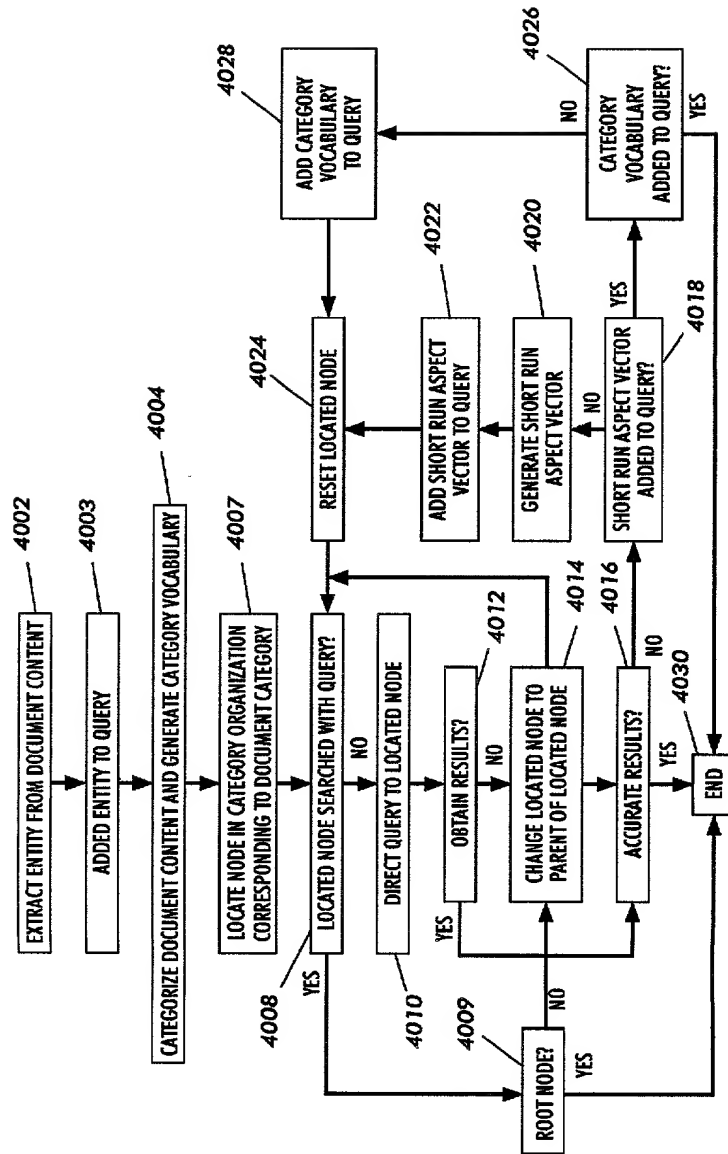
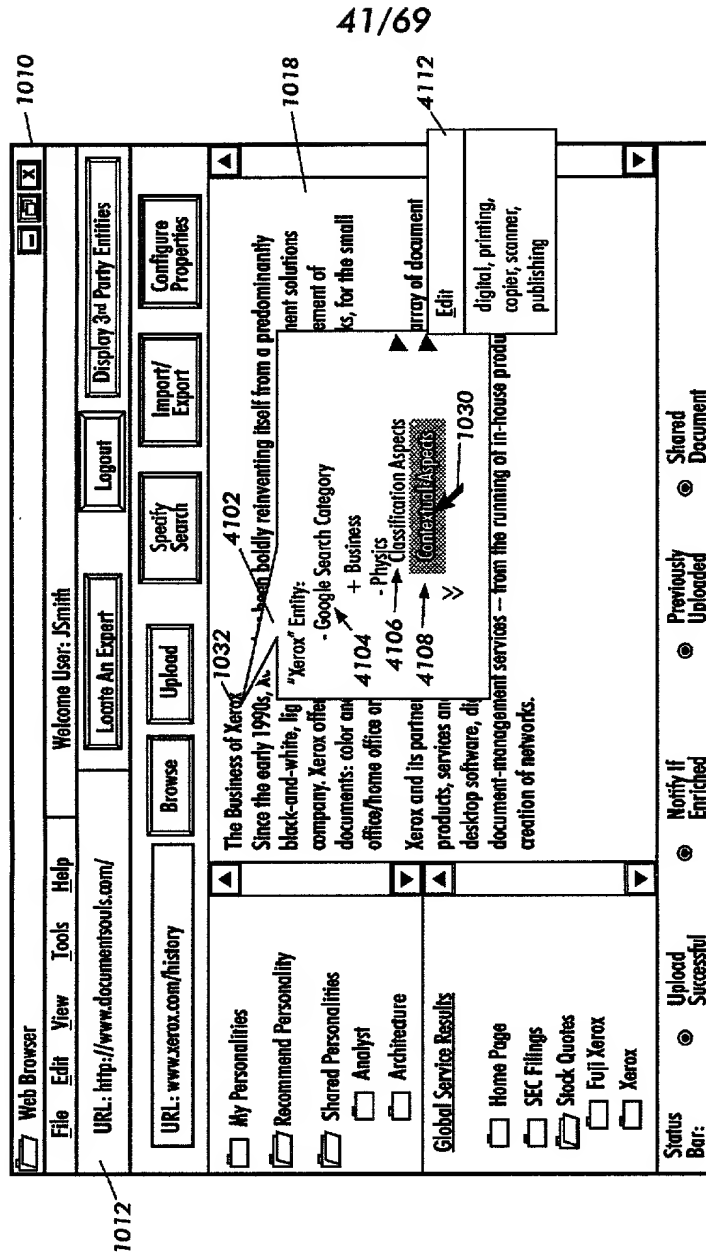
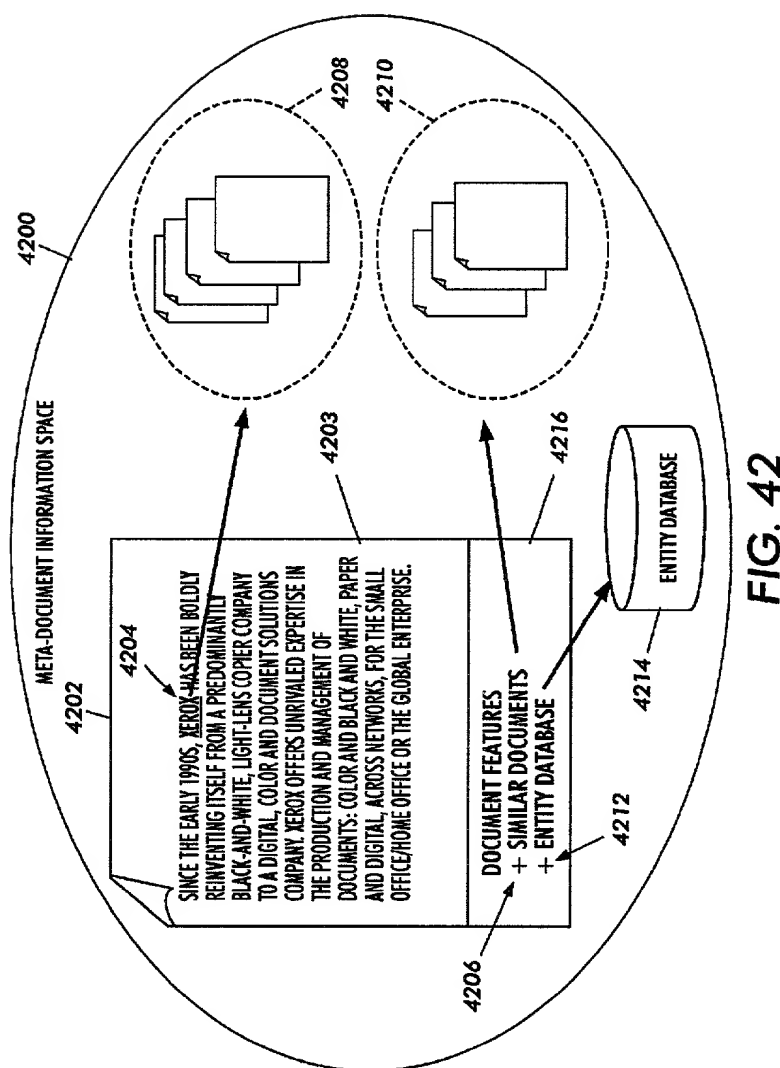


FIG. 40





42/69



43/69

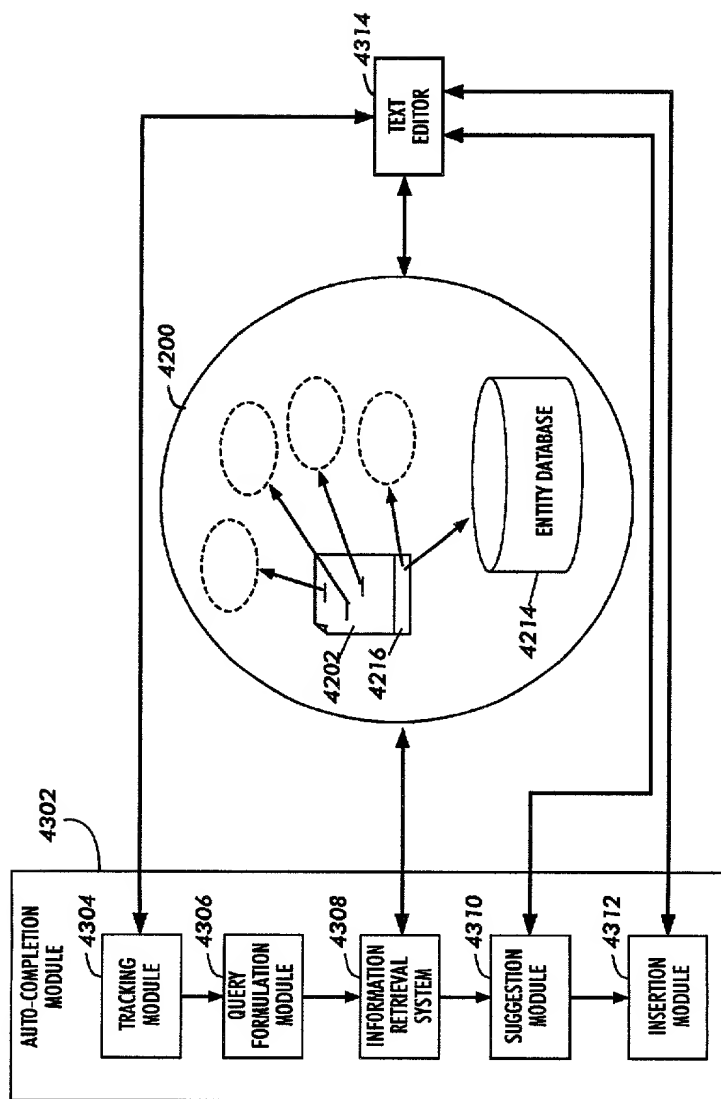


FIG. 43

44/67

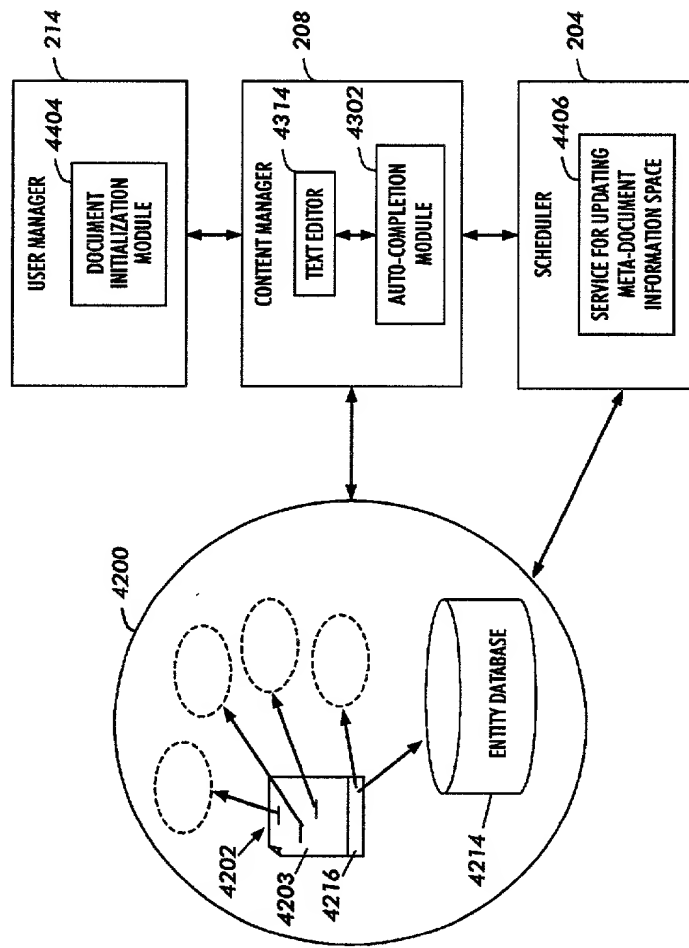


FIG. 44

45/67

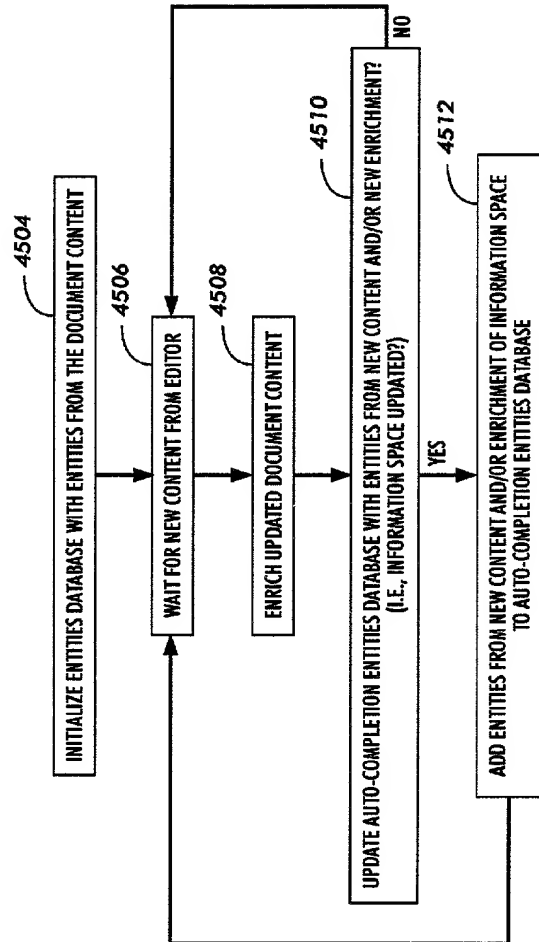


FIG. 45

46/67

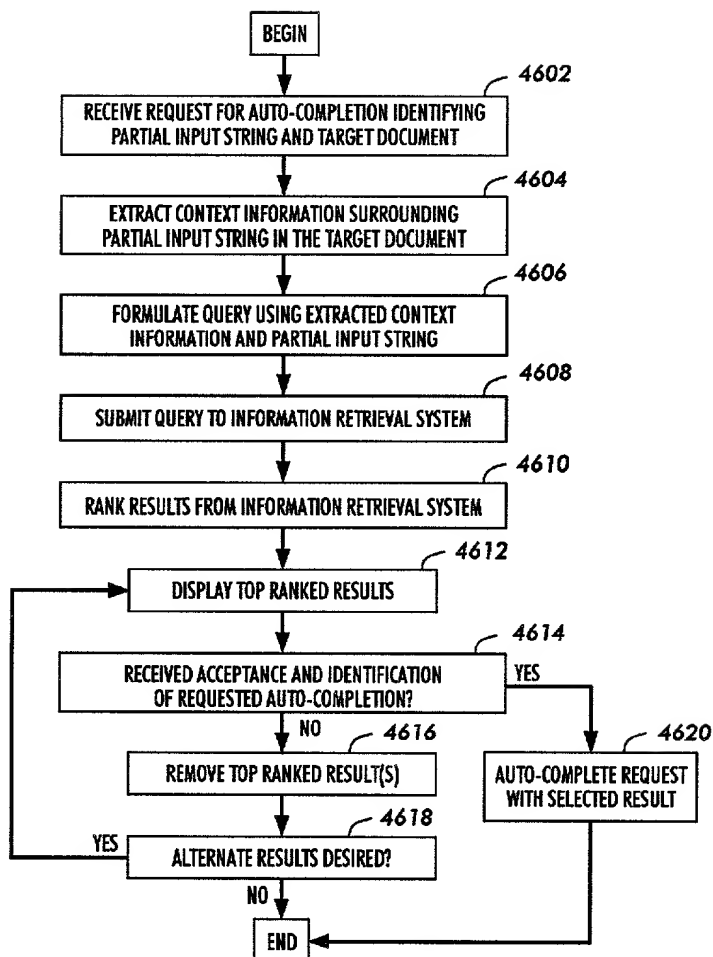
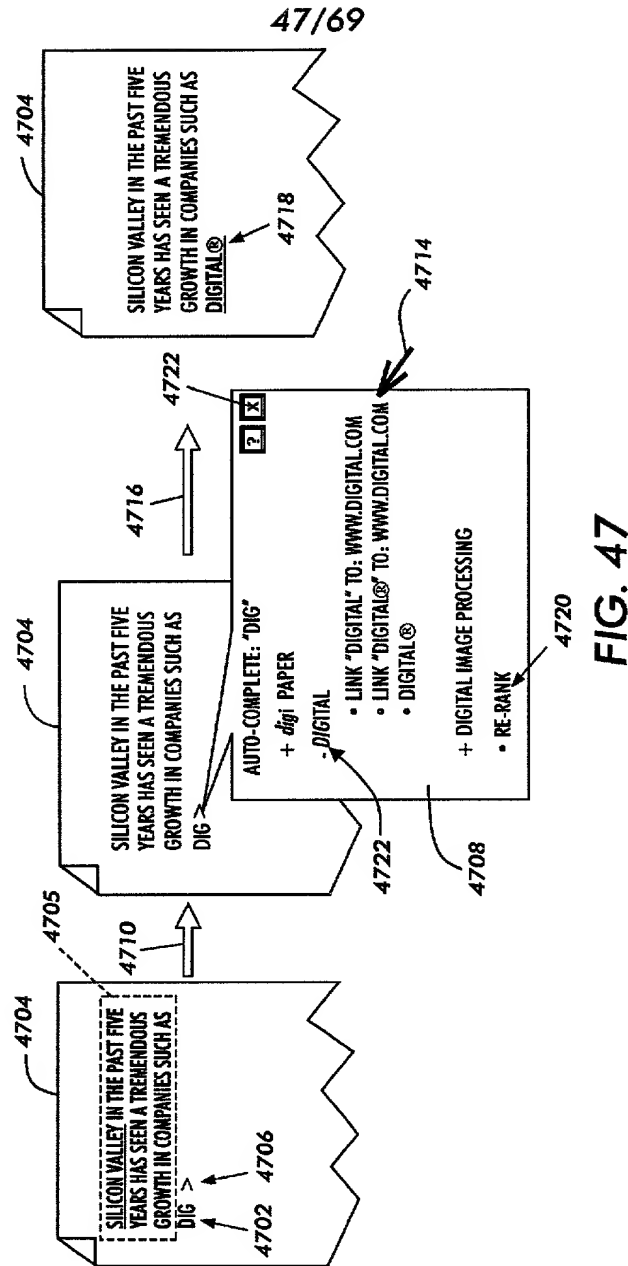


FIG. 46



48/69

(A) KEY	(B) ENTITY	(C) MARKUP	(D) FORMAT	(E) ORIGIN	(F) ORIGIN POSITION	(G) PART OF SPEECH	(H) CONTEXT
Sili	Silicon Valley	www.siliconvalley.com	Capitalize	TARGET DOCUMENT CONTENT	1	Noun	Science
Dig	Digital	www.digital.com	®	www.digital.com	434	Adjective	Science
Dig	DigiPaper	NONE	Bold	www.parc.xerox.com	23	Noun	Science
Dig	digress	NONE	NONE	www.nytimes.com	343	Verb	Politics
...	...	...	...	...	...	...	...

FIG. 48



49/69

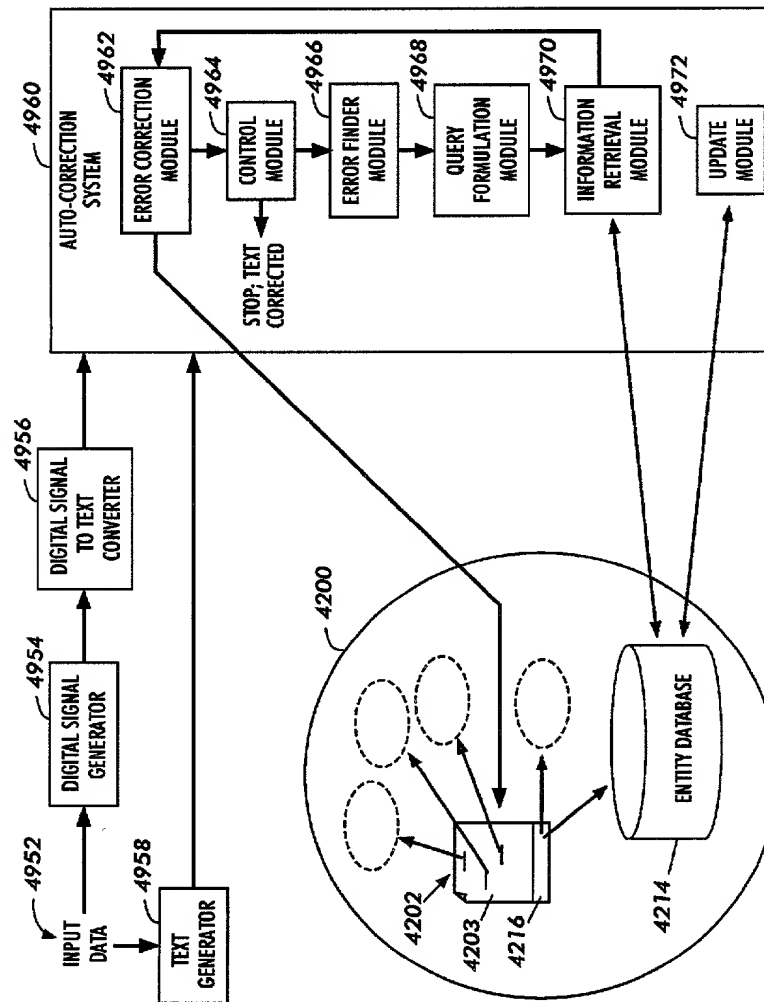


FIG. 49

50/69

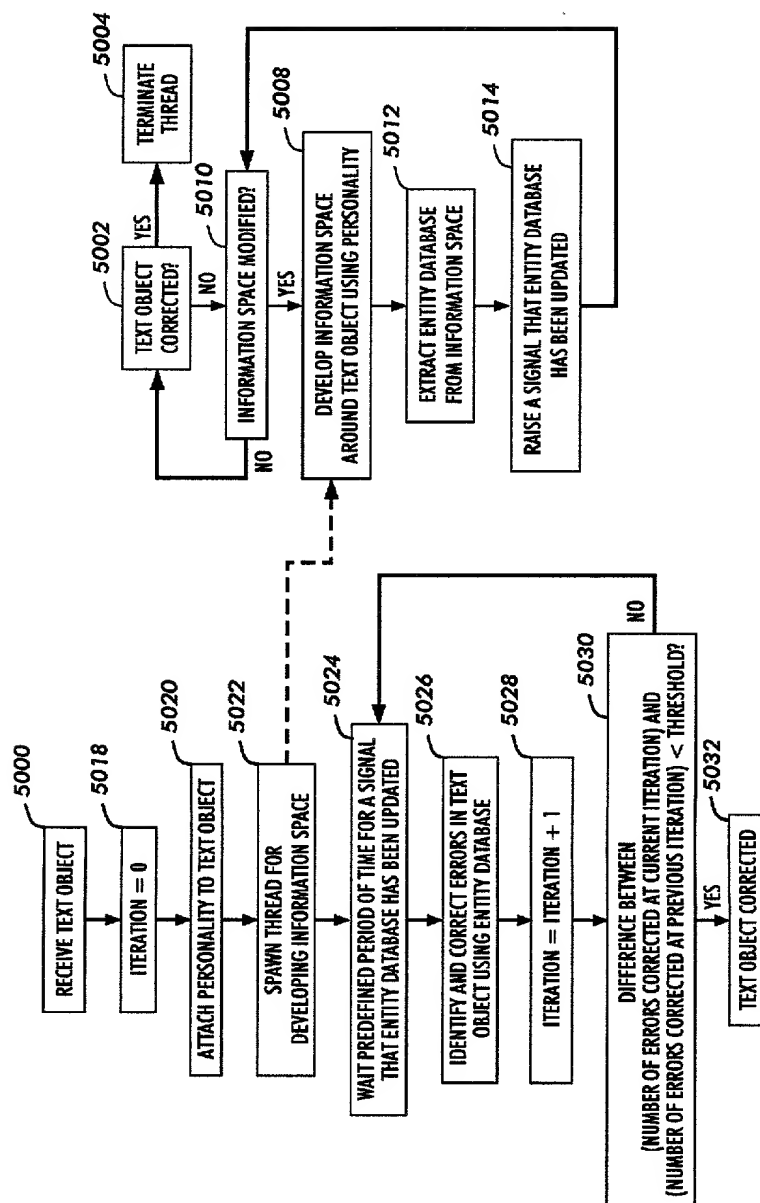


FIG. 50

51/69

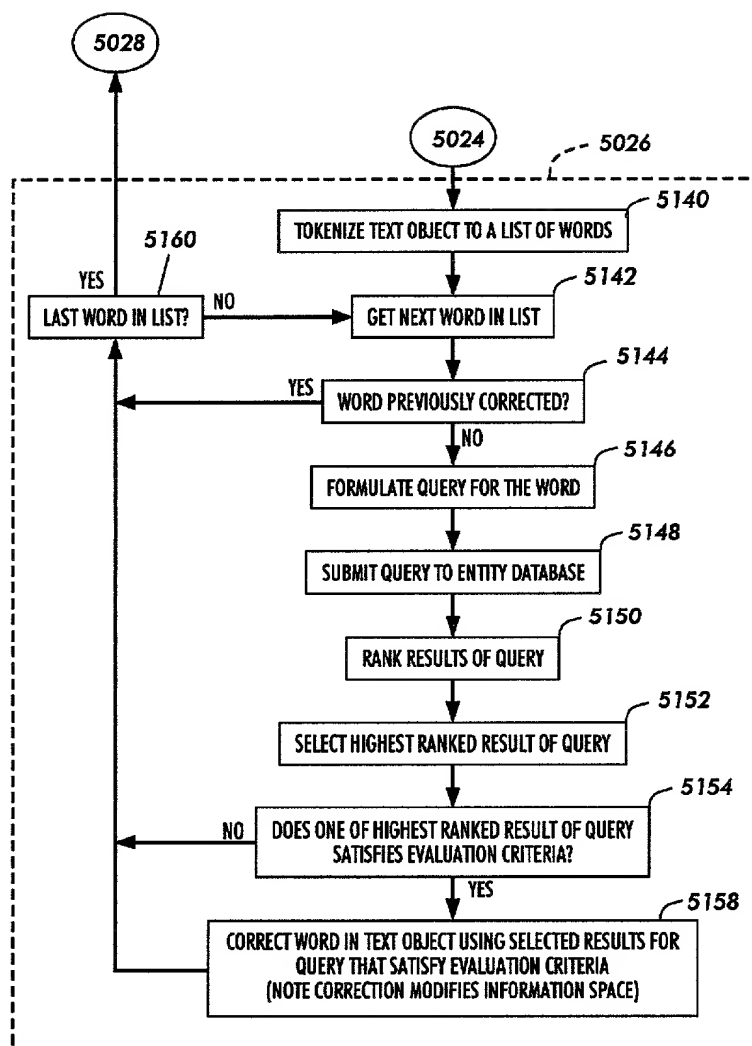


FIG. 51

52/69

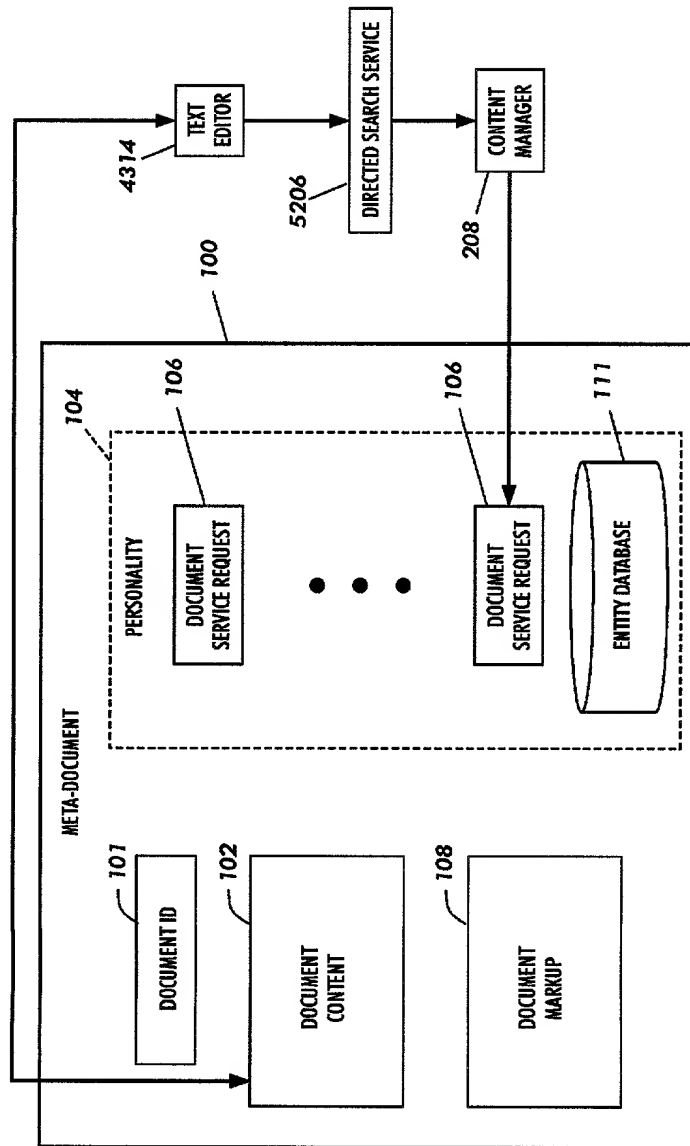


FIG. 52

53/69

**5300** DIRECTED SEARCH **5304**

**5302** Get: Products For: Xerox

**5306** Insert Result as ☐ Link ☒ Content ☐ Metadata

**5308** Update Frequency: Monthly

**5310** Select Providers

- Reuters
- Thomson
- Hoover

**5312** Ensures:

Language of text is: English

☒ Premium ☐ Non-Premium

☐ Low Quality ☒ High Quality

☐ Copyrighted ☒ Not Copyrighted

☐ Free ☒ Paid Max price \$: 50

Maximum number of words: 2000

**5318** Search Selected Providers By:

**5320** ☒ Cost

**5322** ☐ Speed

**5314** Location Of Results:

☒ After Current Selection

☐ Before Current Selection

☐ As Comment To Current Selection

☐ As Footnote To Current Selection

**5315** Current Selection Is:

Xerox in the past five years has seen tremendous growth

**5317**

**5316** Cancel Insert

FIG. 53

54/69

5400

```
<service id='Products'
  refresh=30
  insert='Content'
  language='En'
  quality='Premium&High&Not-copyrighted&paid'
  maxprice='50'
  maxwords='2000'
  search='cost'>
  <entity>Xerox
    <offsets>13 18 456 461 710 715</offsets>
    <provider name='Thomson' id=18></provider>
    <provider name='Reuters' id=19></provider>
  </entity>
</service>
```

FIG. 54

55/69

5500

**IMPORT/EXPORT** [?] [X]

EXPORT NAME:   5502

**EXPORT:**

- ☒ ORIGINAL DOCUMENT CONTENT
- ☒ DOCUMENT MARKUP (I.E., ENRICHMENT)
- ☒ PERSONALITIES

5504

5506

5508

☐ IMPORT NAME:   5512

**IMPORT:**

- ☒ ALL CONTENT
- ☐ SELECTED CONTENT ONLY
  - ☐ ORIGINAL DOCUMENT CONTENT
  - ☐ DOCUMENT MARKUP (I.E., ENRICHMENT)
  - ☐ PERSONALITIES

5510

- ☒ BIND UN-MATCHED ELEMENTS

5514

- ☒ ENHANCE BOUND ELEMENTS WITH EXISTING SERVICE PROVIDERS

5514

FIG. 55

56/69

```

<!-- SERVER ID>
<docsouls:serverid> DS v1.0 server name </docsouls:serverid>
<!-- PROPERTIES OF DOCUMENT>
<docsouls:docid> xxx </docsouls:docid>
<docsouls:name> xxx </docsouls:name>
<docsouls:owner> xxx </docsouls:owner>
...
<!-- ORIGINAL DOCUMENTS DOCUMENT CONTENT>
<docsouls:originaldoc format="HTML"> toto.html</docsouls:originaldoc>
<docsouls:browsingdoc format="HTML"> browse.html</docsouls:browsingdoc>
...
<!-- LIST OF PERSONALITIES AND DOCUMENT SERVICE REQUESTS>
<docsouls:endowmentdescription>
<!-- LIST OF PERSONALITIES>
<docsouls:personality id = " " name=" " etc.>
<docsouls:service id=" " name=" " etc.>
<docsouls:provider name=" " etc.>
<docsouls:provider name=" " etc.>
</docsouls:service>
<docsouls:personality>
...
<!-- LIST OF SERVICE REQUESTS>
<docsouls:service id=" " name=" " etc.>
<docsouls:provider name=" " etc.>
<docsouls:provider name=" " etc.>
</docsouls:service>
...
</docsouls:endowmentdescription>

```

FIG. 56



57/69

```

<!-- DESCRIPTIONS OF PERSONALITIES>
<docsouls:personalitydescription>
  <docsouls:id> xx </docsouls:id>
  <!-- all the fields AND service categories>
  <docsouls:servicecategory id="" name = "">
</docsouls:personalitydescription>

<!-- NEXT PERSONALITY DESCRIPTION>
...

<!-- DESCRIPTION OF DOCUMENT SERVICE REQUESTS>
<docsouls:servicedescription>
  <docsouls:id> xx </docsouls:id>
  <!-- list all fields (e.g., "refresh period", category, etc.)>
  <!-- list all providers>
  <docsouls:provider id="" name = "">
  <!-- list all strategies>
  <!-- list status of services (original, frozen, mapped, etc.)>
  <docsouls:servicestatus status="mapped" originalid="" originalserver="">
  <!-- list ServerId if the service was not available on current server>
  <docsouls:serverid>id</docsouls:serverid>
</docsouls:servicedescription>

<!-- NEXT SERVICE REQUESTS DESCRIPTION>
...

```

FIG. 57

58/69

```

<!-- DESCRIPTION OF PROVIDERS >
<docsouls:providerdescription>
  <docsouls:id> xx </docsouls:id>
  <!-- all the fields e.g. key, URL, etc.>
</docsouls:providerdescription >

<!-- DESCRIPTION OF DICTIONARIES >
<docsouls:dictionarydescription>
  <docsouls:id> xx </docsouls:id>
  <!-- all the fields, e.g. category>
</docsouls:dictionarydescription>

<!-- DESCRIPTION OF STRATEGY >
<docsouls:strategydescription>
  <docsouls:id> xx </docsouls:id>
  <!--all the fields>
  <!--all the dictionaries>
</docsouls:strategydescription>

<!-- RESULTS OF GLOBAL SERVICES >
<docsouls:globalservice id="Summary">
  <docsouls:link> xxx </docsouls:link>
  <docsouls:content> xxx </docsouls:content>
</docsouls:globalservice>

<!-- OTHER RESULTS (E.G. KEYWORDS, CATEGORIES, ETC.) >

```

FIG. 58A

59/69

```

<!-- (OPTIONAL) DESCRIPTION OF ENRICHMENT PERFORMED AT OTHER DOCUMENT ENRICHMENT
SYSTEMS>
<docsouls:serverid>olderServerId</docsouls:serverid>
<docsouls:personalitydescription>...</docsouls:personalitydescription>
<docsouls:servicedescription>...</docsouls:servicedescription>

<!-- (OPTIONAL) LIST OF RESULTS OF THE LAST ENRICHMENT>
<!-- IDENTIFY DOCUMENT SECTION IN CASE OF MULTI-PART DOC>
<docsouls:docpart>main frame.html</docsouls:docpart>
<!-- OFFSET WHEN ENRICHMENT APPLIES TO A SECTION>
<docsouls:offset> 0 200 </docsouls:offset>
<!-- STATUS OF SERVICE APPLIED ("ACTIVE" OR "FROZEN")>
<docsouls:service id="StockQuote" status="active">
  <!-- ENTITY>
    <docsouls:entity>Xerox
      <!-- NORMALIZED FORM>
      <docsouls:normform>Xerox</docsouls:normform>
      <!-- LIST OF OFFSETS WHERE ENTITY IS FOUND IN TEXT RELATIVE TO SECTION>
      <docsouls:offsets start=35 end=40 />
      <docsouls:offsets start=232 end=237 />
      <!-- DICTIONARIES>
      <docsouls:dict> xxx </docsouls:dict>
      <!-- STRATEGIES>
      <docsouls:strat> xxx </docsouls:strat>
      <!-- LINK INFORMATION>
      <docsouls:link> xxx </docsouls:link>
      <!-- CONTENT INFORMATION>
      <docsouls:content> xxx </docsouls:content>
    </docsouls:entity>
    <!-- NEXT ENTITY>
    ...
  </docsouls:service>

```

FIG. 58B

60/69

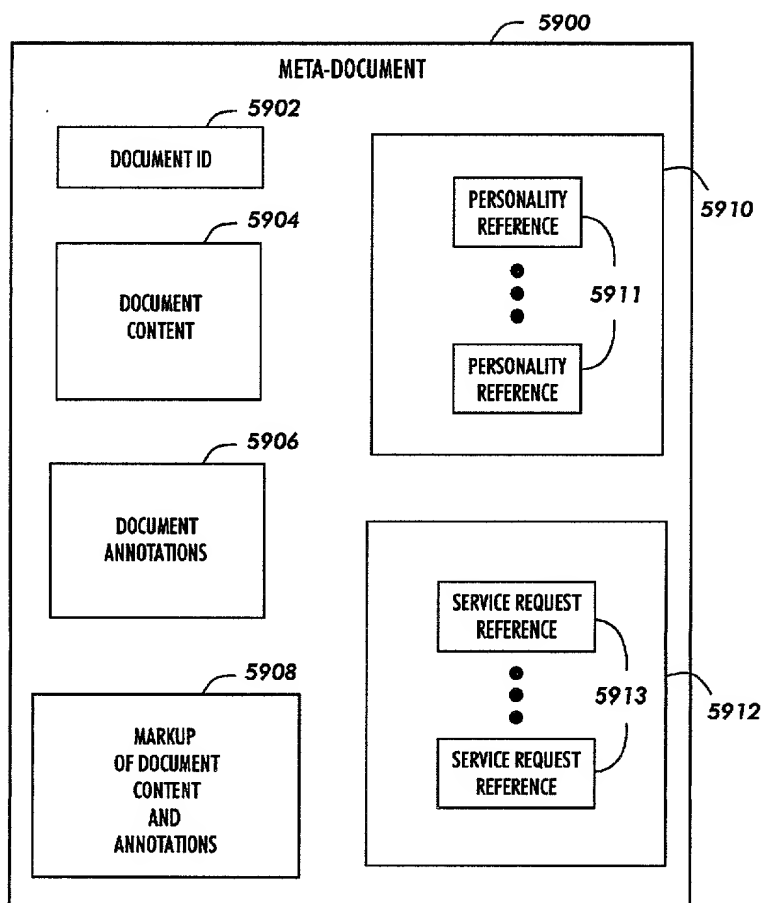


FIG. 59

61/69

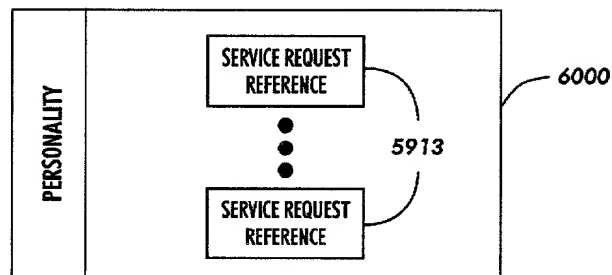


FIG. 60

62/69

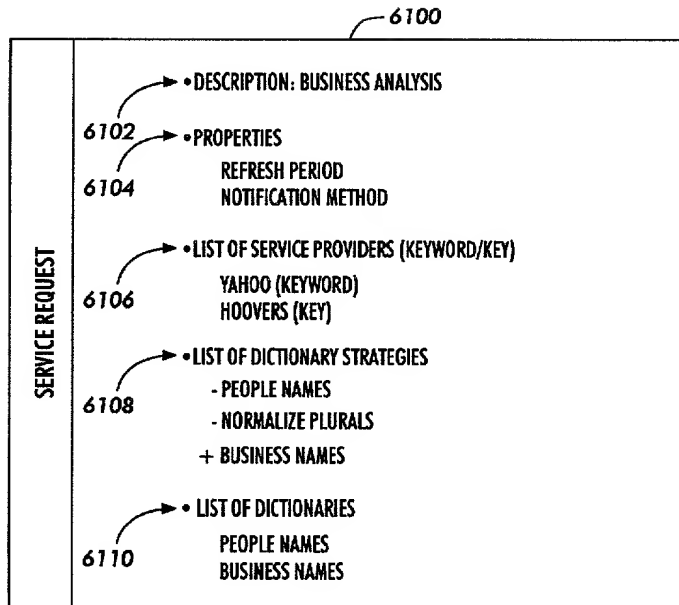
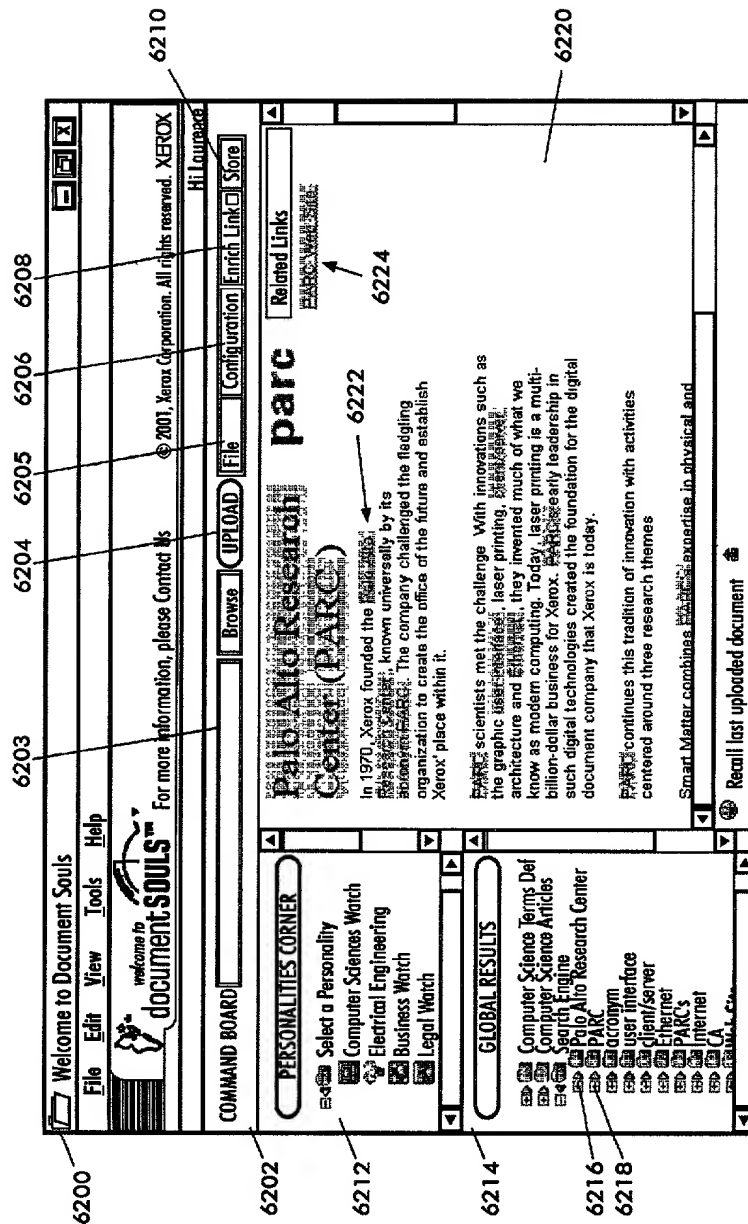


FIG. 61

63/69



64/69

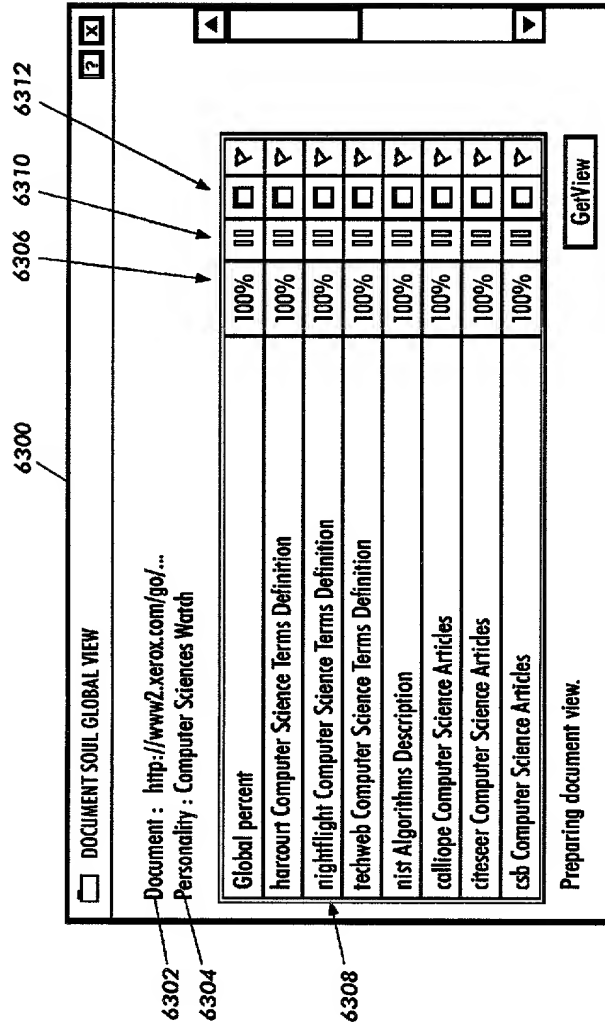


FIG. 63



65/69

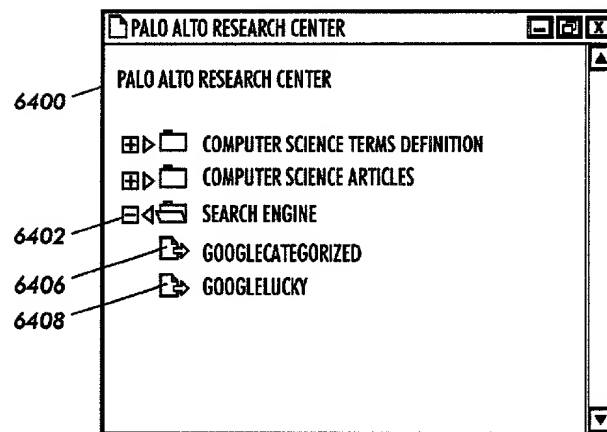


FIG. 64

66/69

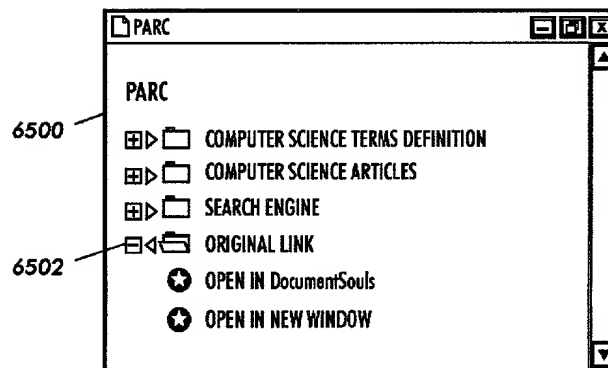


FIG. 65

67/69

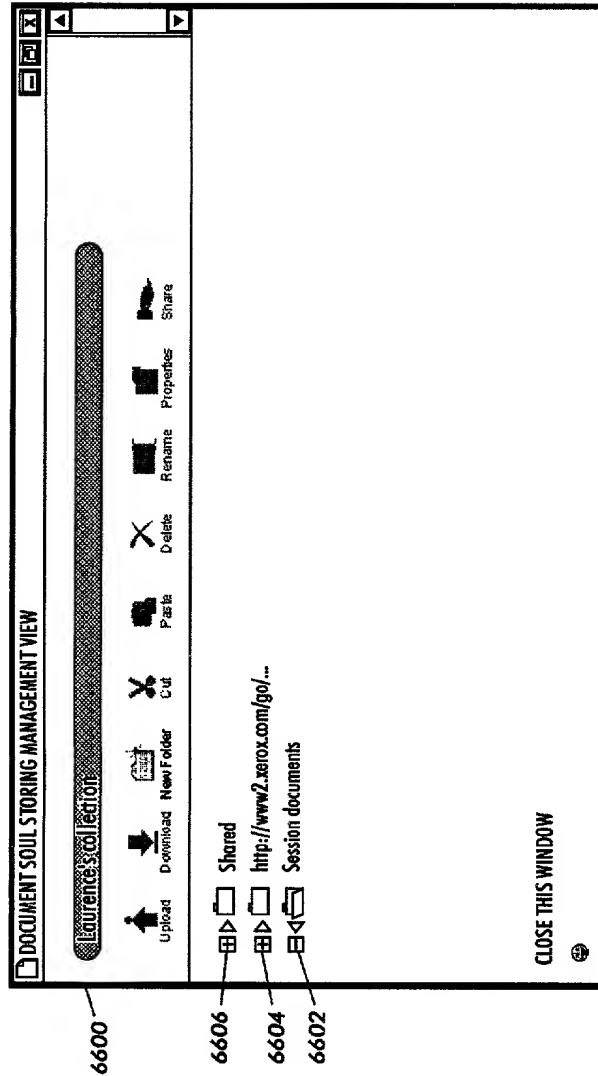


FIG. 66

68/69

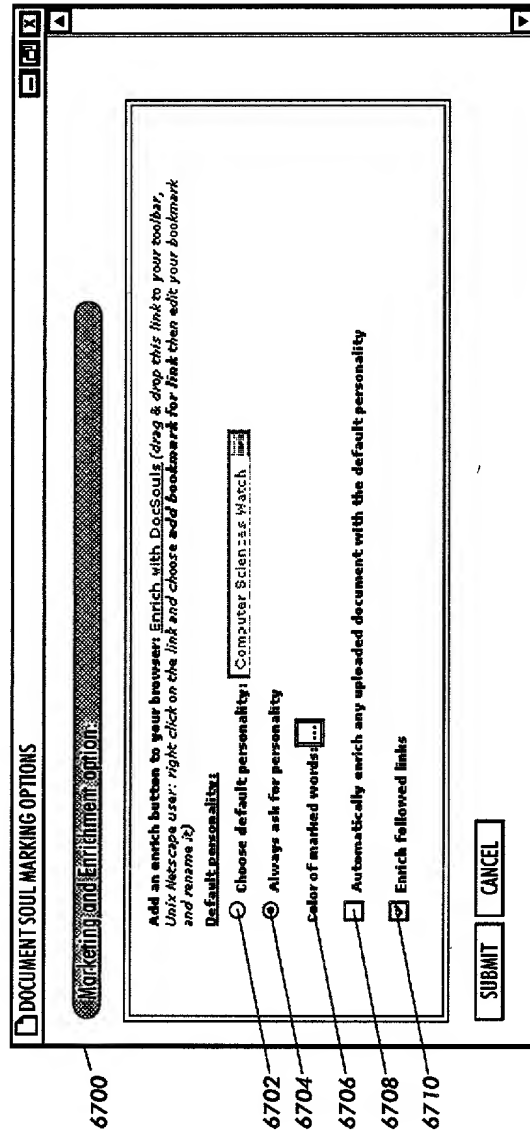


FIG. 67

69/69

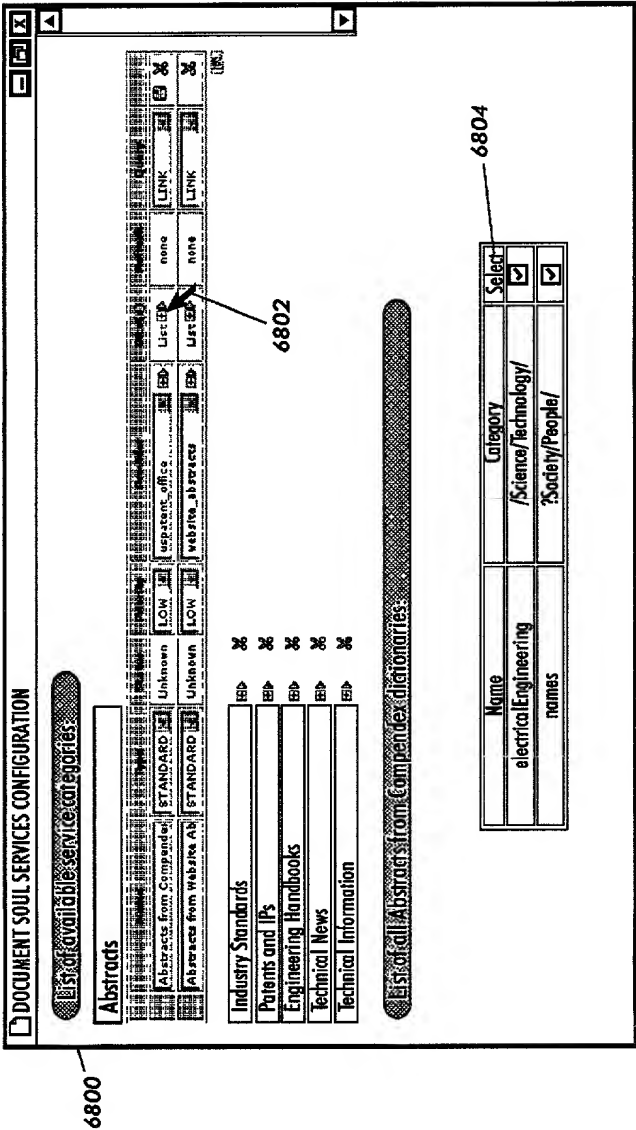


FIG. 68